KUNGL
TEKNISKA
HÖGSKOLAN

SICS
Swedish Institute of Computer Science

Toward Human-Robot Collaboration

Kristian T. Simsarian

Dissertation, March 2000
Computational Vision and Active Perception Laboratory (CVAP)

Akademisk avhandling för
teknisk doktorsexamen vid
Kungl Tekniska Högskolan

March 2000

NADA, KTH, 100 44 Stockholm

ISRN KTH/NA/P--00/04--SE
ISSN 1101-2250: TRITA-NA-P00/04
ISBN 91-7170-544-9

SICS, Box 1263, 164 29 Kista
ISSN 1101-1335: ISRN SICS-D–28–SE

KTH Högskoletryckeriet
Stockholm 2000

Abstract

Recently robots have been launched as tour-guides in museums, lawnmowers, in-home vacuum cleaners, and as remotely operated machines in so-called distant, dangerous and dirty applications. While the methods to endow robots with a degree of autonomy have been a strong research focus, the methods for human-machine control have not been given as much attention. As autonomous robots become more ubiquitous, the methods we use to communicate task specification to them become more crucial. This thesis presents a methodology and a system for the supervisory collaborative control of a remote semi-autonomous mobile robot. The presentation centers on three main aspects of the work and offers a description of the system and the motivations behind the design. The supervisory system for human specification of robot tasks is based on a Collaborative Virtual Environment (CVE) which provides an effective framework for scalable robot autonomy, interaction and environment visualization. The system affords the specification of deictic commands to the semi-autonomous robot via the spatial CVE interface. Spatial commands can be specified in a manner that takes into account some specific everyday notions of collaborative task activity. Environment visualization of the remote environment is accomplished by combining the virtual model of the remote environment with video from the robot camera. Finally the system underwent a study with users that explored design and interaction issues within the context of performing a remote search task. Examples of study issues center on the presentation of the CVE, understanding robot competence, presence, control and interaction. One goal of the system presented in the thesis is to provide a direction in human-machine interaction from a form of direct control to an instance of human-machine collaboration.

Acknowledgments

This work has been a long time in the making. This list is representative of that time and of the people I met along the way. For all their help in direct and indirect ways I wish to express my sincere appreciation to the following people:

To *Jan-Olof Eklundh* and later *Henrik Christensen* who enabled me to perform this work, both by providing guidance and resources. Without these two, and the final deadlines, these words would not be on this page right now. Their patience and rigor has been invaluable over the years. Their contribution here is manifest.

To *Tom Olson*, my first advisor and *N. "Nandhu" Nandhukumar* who lead me through some rigorous research that became my Master's thesis and to *Chris Koeritz* who supplied the office stimulation and distraction necessary to jump that first hurdle.

To *Maja Matrić* for being a research colleague and friend, from whom I drew much research inspiration, and from whom I learned to refine my research writing skills. To *Rod Brooks* without whose example and encouragement I might have left robotics research years ago. To *Luc Steels* for encouraging me to pursue ideas about robot behaviors and human interaction. *Tom Mitchell* for being encouraging and for posing the "fetching" challenge to robotics researchers, and in general to the organizers of the NATO ASI Biology and Technology of Autonomous Agents in 1993. It was that event in 1992 that changed the course my life took for the next eight years.

To *Michael Lebowitz, Peter Allen,* and *Julian Hochberg* at Columbia University who started me on my research career with research assistantships and projects that taught me, by example, that research or researchers do not have to be stuffy or boring. To *Mike Gorman, Bernie Carlson,* and the "Repo research team" at UVa, who showed me that researchers can be nutty, in a fun way. To *Peter Hindle* and *Mr. Haas,* my high school and elementary school Mathematics teachers who's influence carries strongly into the work I do today. To *Bryce Lambert,* my high-school English teacher, who took his knife and cut the word "very" from my essays and taught me that "interesting" was a meaningless word - your voice has been in my head as I wrote this.

To *Amber Settle* and *Mark Fasciano* who supplied me with good friendship, good food and sane voices when things were looking bleak. I thank them for their direct and indirect help in finding a healthy direction for myself and my research. To *Barry Guiduli* who has been a good and constant friend in always challenging me intellectually and physically.

To *Anne Smith,* who encouraged me to move ahead, not just in words, but by example. To my friends *Anne Wright* and *Suelette Dreyfus,* both of whom now have Ph.D.s, and who have tried to challenge, poke, probe, goad and encourage me into finishing - I thank you for that. To *Robert Eklund,* who has provided much intellectual stimulation over the years and who has not yet finished his Ph.D., and to whom I hope to provide similar example.

To *Carl Feynman* who showed me how to break problems down and perform mid-tech experiments, resulting in my learning how simple research can be. To *Chris Csikszentmihalyi* for his technical and intellectual stimulation, artful inspirations over the years, and friendship, helping my work to be better and more orthogonal than it might have been otherwise. To *Ben Bederson and Allison Druin* who have supplied commentary and encouragement throughout the years and who together with *Susan Ettner* have urged me to "hurry up and finish."

To *Jan Seide* for discussions about relativism, reductionism and functionalism and for providing references over the years to the debates in physics and science (and for all the music). To *Julio Fernandez* for being a good co-teacher in behavioral robotics and for providing me with an enriching teaching experience in Spain. To *Phil Agre* for being the intellectual that he is, breaking ground that helped me to realize there are many paths one can take, and that one should trust one's own intuition.

To *Lennart Fahlén* who has tried his best to help me finish and provide me with resources despite competing interests and to *Janusz Launberg,* who has been active in finding resources to enable this work.

To *John Bowers* for providing an intellectual mentorship, an example of someone who can intellectually and rigorously pursue what they enjoy and be darn good at it as well as for reading thesis drafts and offering comments that provoke and validate. To *Yngve Sundblad, Tom Rodden,* and *Jon O'Brien* for reading versions of my thesis and providing me with commentary and support I might not otherwise receive within the CSCW discipline. Also I thank these three for giving me the encouragement to finish.

To *Tomas Uhlin,* the only person I knew before coming to Sweden, who encouraged me to start working on my Ph.D. here and who provided encouragement throughout the years. To *Lars Bretzner and Peter Nordlund* who have provided friendship and examples of good research work, and especially Lars for helping guide me through the bureaucratic process of finishing. To all the people at CVAP-CAS who have helped in seemingly small ways over the years: letting me in the door, helping to carry the robot, lending me a cable, or some advice - that help was invaluable.

To *Martin Nilsson,* who originally brought me to Sweden, and who has directly and indirectly supported this work. Without Martin, this thesis would certainly not exist, I would not have spent time in Sweden, and I would have not had this fine robot to do experiments on. To *Per Kreuger* who introduced me to Yoga and provided a generally positive orthogonal influence, and to *Alf* who tried to help in his own way.

To *Sebastian Thrun* for his advice and for his work with the Rhino system which became the low-level software platform for the robot in these experiments. To *Jussi Karlgren, Ivan Bretan, et al.* for their work on the DIVERSE project and for providing stimulating discussions over the years. To *Tomas Axling, Scott Mcglashan* for their work on the next version of the speech system and to Tomas for his help in constructing an interface to the robot.

To the administration at SICS for helping with all details along the way and especially to *Lotta Jörsätter* who was able to get the robot in and out of the country a number of times, and to *Marianne Rosenqvist* who has helped at a number of important times.

To *Kristina Höök, Lars Oestreicher, and Kerstin Severinson-Eklundh* for helping to form the ideas that became the Study of Use in this thesis. As well as to all the colorful people who participated in the study, without their vivid and candid remarks there would not have been much to write about in chapters 5 and 6.

To the entire *ICE lab* for all their help and support over the years. Certainly without DIVE this system would not have been built. To *Karl-Peter Åkesson,* especially, for his exjobb on the Reality Portal system in this system and his co-authorship on a series of papers upon which some of this work is based. To *Emmanuel Frécon* for helping to construct the first version of the robot-DIVE communication channel and

for some of the excellent images created for papers, a few of them re-appearing in this thesis. To *Olov Ståhl* for constructing and working on the video stream within DIVE. To *Marten Stenius and Daniel Adler* for helping in the CeBit demonstration in 1996. As well as to *Par Hansson and Anders Walberg* for doing fine work in the Grotto with user interfaces and helping to improve a number of DIVE features used in the construction of this system. To *Lasse Nilsander and Peter Elin* for being responsive to technical problems encountered along the way.

To *Dodi* whose energy and help were valuable. To *Åsa Rex* who provided good conversation and helped to make my initial time here a lot more fun.

To *Jerry Garcia, Robert Johnson* and other 'friends' as well as to the *Ashland and Larrivee* guitar makers for providing inspiration and a respite from the technical.

As is the way when one does their best work, many of the above are both friends and colleagues. Even more significant is the contribution of my family. To *J.R. Simsarian, my father*, for awakening my technical interests at a very young age and for helping me out financially when I needed it during my educational journey. To *Carol and Gordon Loughlin* for providing me an asylum from all the craziness, and giving me love, sunshine and happy times.

Most of all, my appreciation and love goes to *my mother, Astrid Tollefsen,* for endowing me with initiative and the quality of not being afraid to work hard and for all her support and love over the years. Certainly I would not be at this stage without her.

Contents

Chapter 1

Introduction

WE are in an age where the number of autonomous machines around us is increasing. Though this has been true for many decades, it is now becoming true with respect to the general citizen. This marks a shift of deployment from the laboratory and the specialized user to the non-expert. This is also true in the sense of who is specifying the robot tasks. Recently we have seen the launch of mobile robots as tour-guides in museums, as lawnmowers, as home vacuum cleaners, as home-care aids, as intelligent play toys, as remotely operated machines through the WWW and generally as telerobots in so-called *distant, dangerous* and *dirty* applications. One can reasonably expect this trend to continue. Many of the capabilities of these autonomous robots can be seen as the direct result of basic research within the robot research community. While the methods to endow robots with some degree of autonomy have been a strong focus of research, the methods for human-machine control have not been given as much attention. Rarely are methods for human-robot interaction the focus of mobile robot research. With this increased presence of robots comes a need for more exploration in the way the interaction takes place between humans and autonomous machines. As autonomous robots become more common, the methods we use to communicate task specification to them become more crucial. If we can assume that in the end it is the machines that will do our bidding, then the methods for the communication of our goals and the methods for promoting user-understanding of a robot and its environment must take a more central role.

For many of these tasks there is a need for the human user to be aware of the robot's remote environment and of the robot's capabilities. Such awareness includes being able to visualize the robot's situation and environment, and once provided with this awareness, having the methods for instantiating commands for a robot to perform. Underlying this awareness and the control methods is an infrastructure framework within which the interaction itself takes place and enables communication of the commands to the robot. This framework

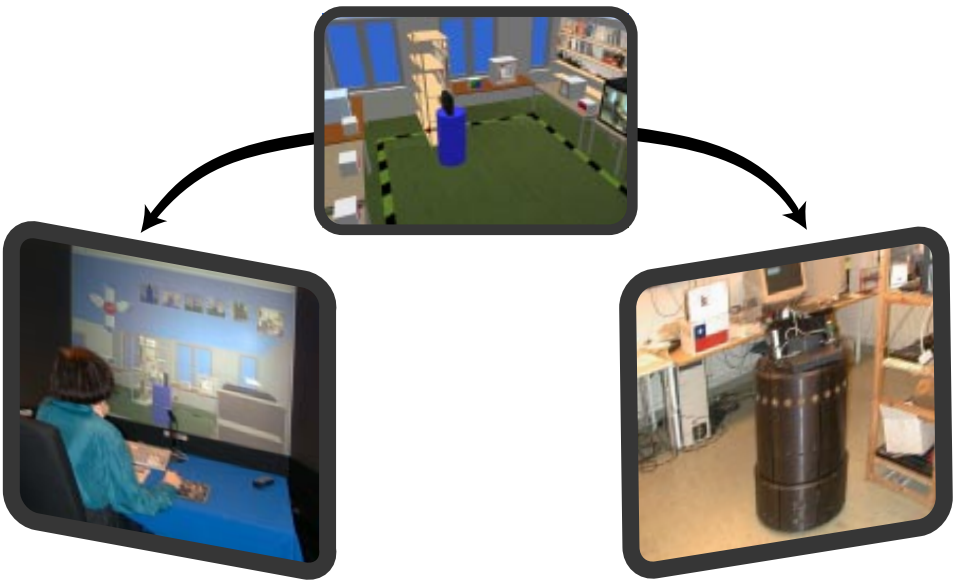

Figure 1.1: This figure shows the three main parts that make up the complete system. The human user, the Collaborative Virtual Environment, the remote robot and also depicts the CVE as the medium of communication between the human and the robot.

generally consists of the metaphors employed and the technical implementation that supports the control and task-specification of the autonomous system.

1.1 Objectives

This thesis looks at some particular solutions which include a shared representation of a robot environment can be provided to a user, the methods for how human supervisory control can be specified by a human supervisor to a robot assistant and the guiding metaphors and structure in which this interaction takes place. In particular, the thesis presents the technical and methodological aspects of constructing a spatial interface for remote robot control and its supporting framework. The framework consists of an infrastructure and a guiding metaphor. The infrastructure is composed of a Collaborative Virtual Environments (CVE) and semi-autonomous robot and the guiding metaphor is that of supervisory control. Under supervisory control, a human operator specifies task-level commands to a remote robot. Different aspects and requirements for such a system are given, and a set of solutions are provided. Specifically,

the methods for robot environment awareness are seen here as methods for *environment visualization* and the methods for control are different interaction mechanisms for communicating task commands to the robot. For the interaction, an approach employing Collaborative Virtual Environments is used in combination with a deictic speech and mouse-based gesture interface. For the visualization of the remote robot working environment, a number of solutions are implemented progressing toward a blend of the real video textures from the remote site and the 3D graphic model of the working environment creating an instance of *Augmented Virtuality*.

Thus, the practical problem in this research is how a human can communicate spatio-temporal task specifications to a robot along with the principles and techniques that help guide this interaction. In response to that problem, this work: identifies the central interface elements of this interaction that need attention, implements and demonstrates the prototypes, and performs a design study with users to look for actionable insights into the construction of human-robot systems.

1.2 Approach

Key problems of interest in human-robot interaction lie in visualization of the remote environment, multi-modal specification of spatio-temporal tasks and in the framework for the robot-human relationship. Work in robot-human communication area seeks to improve robot utility, ease of expressing tasks, the flexibility of the communication, and to generally better the understanding of how to build robot-human interfaces.

The guiding metaphor for interaction between the operator and robot is to consider the robot as an *assistant*. It is this concept that drives the conceptualization of a semi-autonomy both within the robot sub-system as well as when viewed from the outside by the operator. The operator and the robot, together, form a supervisor-assistant relationship and it is through this partnership they form a collaboration with respect to a given task. The target category of tasks for this system are 'point-to-point' navigation, 'go-and-look' search and 'pick-and-place' manipulation. Although the system comprises several sub-systems, the desired effect is that the system be viewed as a medium for human-machine interaction where the details of the sub-systems unify to implement a system centered around the activity of task specification and interaction. That is to say, interaction is with the task, not just the robot.

The user interface enables communication between the robot and the human operator. This is done through visual and spatial interaction, complemented by gesture and speech. The framework for the interface is based on a Collaborative Virtual Environment (CVE) that provides a medium for spatial interaction between the robot and the operator. The virtual environment is used for task specification as well as world modeling. The tools for interaction combine 3D

gestural commands (e.g. via mouse or other device) as well as speech as the main input between the robot and the human. Most tasks specifications are spatial and take advantage of spatial nature of the 3D interface and robot environment. One of the underlying theories in this work is that within the 3D environment spatial commands can be specified in a way that builds on specific everyday notions of collaborative communication and task specification.

A goal of exploring different techniques for environment visualization is to provide the operator with the ability to explore the model and video of the remote environment without the constraint of robot movement limitations (e.g. from direct manipulation of the robot) or from being limited by the camera field of view. These constitute a set of spatial-temporal problems which find partial solutions in the work here. These solutions are methods for storing the remote images in the 3D model. Specifically, an augmented 3D virtual environment is created which contains real world images as object textures and allows the operator to explore a virtual representation of the real space. The advantage of using this 3D graphical environment is that it can be manipulated in a way not subject to the temporal, spatial, and physical constraints of the real world. It also has the advantage that selected parts of the real world that may be irrelevant to a particular task can be omitted from the rendering. Thus the parts of the world of interest can be extracted and made salient. A subjective video-augmented graphical view of the robot environment can be created, allowing the operator to work with the elements of the robot environment that are relevant to the current task. Different methods for this visualization are explored and presented in this thesis and then explored further in the design study.

1.3 Contributions of this Thesis

The main contribution of this thesis is embodied in the form of a system implementation. In addition to the components that compose this implementation is a *study of use* of the system with a number of users. These contributions can be summarized as the following:

- Framework for Human-Robot collaboration;
- Techniques for remote environment visualization;
- Human-robot multimodal deictic task specification;
- Robot semi-autonomy;
- Demonstrations and a study of use.

These contributions are elaborated below.

Framework for Human-Robot Collaboration

The fundamental element of the technical work in this thesis is the framework for the human-robot collaboration system. This is an infrastructure consisting of a Collaborative Virtual Environment and three-dimensional model of the remote environment, a remote robot, a set of communication and display sub-systems and the physical set-up of the environment of use.

Techniques for Remote Environment Visualization

A significant element of a human-robot system is the ability to visualize the remote environment. This ability is embodied by a set of systems that allow remotely sensed information, e.g. video, to be displayed in the environment. This thesis reports the different methods of visualization that have been employed in the system, how they have been used and how they are constructed.

Visualization of the remote environment is accomplished by unifying the graphical model of the robot environment with video from the robot camera. The system combines 3D graphical environments with live video and images generated from the robot exploration of the remote location. Different techniques for this combination are presented. These techniques differ in their method of incorporating the remote robot video into the 3D environment. The technical basis for the visualization methods is grounded in Augmented Reality (AR) and Augmented Virtuality (AV). This visualization of the remote environment forms an important part of the human-robot interaction.

Human-robot Multimodal Deictic Task Specification

The human-robot collaborative framework provides mechanisms for pointing and interacting with the three-dimensional environment. The system for pointing in the world has been augmented with a speech control system that together enables deictic task specification. The speech and pointing system contain a grammar of commands and a model of interaction that enable the reference to objects in the environment, and transitively, the real world. Because the three-dimensional model is representative of the real world environment, selecting objects in the graphical environment can also be seen as a method for selecting objects in the remote robot environment.

Robot semi-autonomy

The system here attempts to move beyond simple human control of a robot by giving the robot basic competence. There is a scale of autonomy. On one end of this scale is full-autonomy where a robot works completely independently. On the other end of this scale is complete operator control, where the robot is simply a remotely operated vehicle. In this system, the concept of semi-autonomy is explored where the robot and the human cooperate to complete a task. The robot has the basic real-world interaction capabilities of point-to-point navigation, of simple obstacle avoidance and of grasping and releasing

objects. These capabilities in combination with real world range sensing, and communications with the other subsystems form the robot sub-system of semi-autonomy. The human then provides commands that employ these capabilities in a sequence to perform the target tasks of fetching.

Demonstrations and Study of Use

Real-world demonstrations are important to both disseminate the research ideas embodied in a system as well as to learn and understand more about the system, its shortcomings, its advantages, its use, its potential users, and possible applications. Reported in this thesis are the findings of these studies and demonstrations. Most significant of these is the recent *Study of Use*. This Study of Use encompassed setting up an environment, task design, and study execution with users employing the system for a particular remote search task. The study collects their reactions and forms a set of implications for system design. The study is primarily composed of the techniques of observation, a survey questionnaire, and a qualitative interview centered on design. In addition to that study, a number of formal and *ad hoc* demonstrations of the system have taken place over the years since the system has been developed.

1.4 Structure of the Thesis

The thesis presentation is organized into the following chapters, **Introduction**, this chapter where the main elements in the thesis are provided along with an outline of the entire thesis. This is followed by a chapter exploring **Related Research** where work that forms the research background is discussed. Following that is the **Approach** chapter where the motivations and system outline are provided. This is followed by the technical presentation in the **Methods** chapter. The **Study of Use** chapter presents the results of an explorative design study on the system and the **Findings** chapter summarizes the study's main results. The final chapter is the **Conclusion** and includes reflections and a summary of the main points.

Chapter 2

Background and Related Work

2.1 Introduction

The primary question driving this research has been how to perform a task at a distance by employing a remote robot. The resulting solution is an integration of different research disciplines. The system combines real-world autonomous mobile robotics with a simulated virtual environment as a medium for high-level control. The main application for such a system is the usage of mobile robots in the broad area of hazardous or inaccessible environment exploration and remote assistance while also considering the emerging area of domestic robots. The work is interdisciplinary by nature and crosses many traditional boundaries. Though this is a strength of the work, it also leaves it vulnerable from the perspective of any one particular field. This chapter is an attempt to provide a structure of the work and place it in context with previous research. The primary research areas are autonomous mobile and telerobot work, Artificial Intelligence, software agents, human machine interfaces, CSCW, and media space interfaces. Much of this work is fundamentally inter-related making this partitioning somewhat unnatural. The treatment here, by necessity, separates out these disciplines while an attempt is made to amend this by making note where work spans and connects traditional research area boundaries. The description will begin at a general level, covering related work in these different research fields, and then move on to the specific work on human robot interfaces and the studies done to evaluate robot interaction systems.

2.2 Genesis of the System

This thesis work has evolved from the author's previous work in robotics, computer vision and artificial intelligence. The topic of mobile robot self-localization was the topic of a masters thesis [98, 104]. That work inspired later research on a robot perception system that integrated a number of computationally inexpensive techniques for object and location recognition [97]. At the NATO Advanced Study Institute where that work was presented, Thomas Mitchell made a call to the robotics community for a fetching system that could "Take X from Y and bring it to Z" [77]. It is partly that call that inspired this work on a human-robot system. Originally the idea was to employ a human supervisor to handle the hard problems in Artificial Intelligence. Through this human and robot integration an integrated functioning system could be created that demonstrates robot capabilities in the context of human guided use. This work was first reported in 1995 [102] and later refined as a proposal for an assistant to persons with disabilities [101]. The system was further expanded with greater facilities for communication and control by integration with a deictic speech control system which was reported in a book collection [103]. More exploration was done on the visualization aspects of the system and these results are described in [100]. Later a catalogue of these visualization metaphors are presented in [99] and the most recent presentation [6]. This thesis is the sum of this work, plus the theoretical underpinnings and the more recent work including a design study with users.

The work has evolved in this direction from a dissatisfaction with the current state of the art, and even research agenda, of many in the AI community. Many of the hardest problems in AI are those at the highest level while many of the lower-level problems are finding solutions in mobile robotics. In the last decade, research concentrating on these low-level problems has created different engineering approaches that brought robotics to the stage of obstacle avoidance and simple navigation. However few results in "top-down" research have progressed to real solutions that enable a mobile robot to make autonomous decisions within a real unstructured setting. Such unsolved high-level problems include deciding which tasks are important to perform next (given virtually unlimited possibilities), or the more specific problem of how to decompose tasks into subtasks. It was this absence of high-level competence that drove this work in the direction of adding a human supervisor. In a supervisor scenario the human becomes responsible for high-level decisions. The robot, then, is responsible for tasks at its given level of competence: *e.g.* avoiding obstacles, maintaining simple trajectories, grasping and employing simple perceptual processing. Finding the appropriate mode of collaboration is a point of adjustment and agreement, tacit or explicit, between the human and the robot system. The system in this thesis is an example of how such a human robot system might be constructed to enable expansion of robot competence while maintaining a constant interface presentation and metaphor.

Specifically the intended domain of tasks for the system is centered on fetching: "Take X from Y and bring it to Z". This task breaks down into locating X, moving to X, identifying Y, acquiring Y, locating Z and moving to Z. Thus the human operator's role is to supply the instantiations of these variables. With a human operator the task for the robot is considerably easier. In simple domains, the locomotion, the identification and the grasping become engineering solutions, yet the system as a whole becomes powerful. As the mobile robot becomes capable of higher level tasks, it is the intention of the system to enable the operator to assume a greater monitoring role. That is, the robot is under the high-level task direction of the human, but the robot can assume more responsibility for the task as its task competence and autonomy increase.

The subsequent sections present related work in the different research domains related to this system.

2.3 Mobile Robotics

There is a definite advantage in being able to send autonomous robots into hazardous and remote environments, *e.g.* space, sea, volcanic, mine exploration, nuclear or chemical hazards. Robots can be built to stand higher environmental tolerance than humans, perform repeatable specialized tasks efficiently and are expendable. To this end, there has been much research on fully autonomous robots that can navigate into an area, perform actions such as taking samples, performing manipulations, and return the information to base controllers without human assistance.[1] Beyond those institutionally funded exploration applications (*e.g.* the Mars Rover [93], or the Dante volcanic explorer [12]), a number of mobile robot applications have been recently launched by commercial companies for use by the average citizen. Such applications include a robot lawnmower [70], a dust vacuum for domestic use [1, 85], a family care assistant [111], a robotic play toy [32], and a museum tour-guide [28]. The presence of robots outside of the laboratory and in use by the non-technical, non-expert operator puts a greater emphasis on how a human controls a robot. Though this thesis work did not originally address questions about robot use for the general citizen, the intention is that some of this research might help inform the enterprise of creating a focus on the interface. Further, by the time of the study presented in chapter 6, one of the foci of the study became the possible use of the robot in a home environment and the study address some of the issues that might be encountered.

[1] Much of the work in this direction is demonstrated every year at the AAAI mobile robot competition.

2.3.1 Applications of autonomous and semi-autonomous systems

What is the context in which a supervisory robot system may find implementation and use? Red Whitaker has referred to the situations where remote mobile robots are useful as **DDD**, *Dirty, Dangerous, and Distant*.[2] These are more specifically dirty in the sense of having rather low appeal to the human worker. Jobs that, by their nature, involve environments and materials that are fundamentally unappealing, *e.g.* repair work in sewers. For the most part, dirty jobs are those that do not offer an attractive working environment.

Dangerous work has elements of personal risk. Such tasks might involve work with hazardous materials, such as toxic or caustic chemicals or radioactive materials such as nuclear waste, or other extreme conditions such as cold or heat, *e.g.* firefighting. A further example is when the nature of the work may create a hazardous environment, such as in structure demolition. Not only is it cumbersome for humans to protect themselves from such hazardous elements, but such protection may impede the efficiency of the task. Moreover the robot construction might be made to be resistant to such elements and thus be unaffected, or in the worst case, the robot is at least more expendable than the its equivalent human worker.

Distant work is carried out on other planets, or even terrestrial environments such as deep sea, deep mining, or high altitude where the practical limits of human control force the situation toward remote control. When such distant work involves inter-planetary distances there is also a significant time delay incurred in any communication. This introduces new problems and may cause a re-working of the control structure. Delays to Mars, for example, are over ten minutes for one-way communication, making direct control of an autonomous mobile robot impractical. Some applications, such as deep sea and volcanic exploration may cross some of the DDD categories. Typical tasks in such environments are exploration, reconnaissance, inspection, and repair. Thus a situation like the 1986 Chernobyl emergency might have benefitted by having robots on hand that could be sent in as remote agents to examine, shovel dirt, lay concrete and update the status inside the radioactive zones in the plant. As a result of that need at Chernobyl and building on experience with Three Mile Island, the company RedZone was formed in 1987 to construct a robot to send into the Chernobyl Reactor unit[3] In 1999 the RedZone Pioneer robot was sent in to perform reconnaissance and build a map of the area using video techniques together with virtual reality [116].

A further application for teleoperated mobile robots is control over scale, *e.g.* micro and macro robots. The category of micro robots broadly includes

[2] Red Whitaker, Discovery Channel program *Robots Rising*, 1997

[3] Soviet soldiers and other workers, referred to as 'biobots,' have been used for tasks in the reactor area since 1986 at great personal health risk. By western standards the environment is considered poisonous.

those that can work in environments at less than human scale. There has been research in what are referred to as "nanorobtics," robots on the scale of a billionth of a meter. Such robots might be sent into the blood stream to clear the clogged arteries of a patient. A less altruistic application might be insect-sized robots used for surveillance. On the large end of the scale are robots capable of performing work that is beyond human physical limits. Simple examples of these are autonomous logging machines, or the large construction or earth moving equipment used in mining or mineral extraction and transportation. Though the boarder between *dirty, dangerous, distant*, and scale is not always distinguishable, such a work categorization supplies a context in which this system can be considered. The system has been tested over distance (10M to 2000KM) and although the system has not been used in all these contexts, the examples provide a direction in which such systems might develop and a practical motivation for their existence in the first place.

Autonomous and Tele-Robot research Research work in the field of telerobotics and that in autonomous mobile robotics is quite separate. Not only is the research often carried out by different principle investigators, but there are often institutional boundaries to collaboration[4]. Some researchers have tried to bridge the gap between autonomous robots and telerobotic work from both directions. From the perspective of telerobotics this has been an attempt to build more powerful supervisory control systems. From the autonomous robotics perspective, this has often been an attempt to build functioning practical systems that employ the autonomous mobile robot research methods as a foundation implementing supervisory control systems. However the community separation (*e.g.* separate conferences, journals, cultures) often means that communication of the latest techniques and results is lacking.

In two examples of integrating the technology of telerobotics and autonomous robotics there is evidence of the differing perspectives. The autonomous robotics group at the Georgia Institute of Technology has taken a schema-based reactive architecture and used this as a base-level for teleoperated control. In their architecture the mobile robot performs simple navigation while the operator's commands can be situated in the system either as another behavior that influences navigation or as a more global process that manipulates system parameters [8]. They have since extended this idea to allow the operator to control group behaviors in a multi-agent environment [9]. Other groups have recognized the need for tele-operators to move away from low-level robot movement control. One effort has created a multi-level architecture for robots to provide higher level navigation functions such as path-planning and obstacle avoidance with operator guidance [24]. One such sophisticated system was

[4]Whereas most intelligent autonomous robot work is conducted in Computer Science, Electrical Engineering, or Artificial Intelligence departments, much teleoperated work is carried out in Mechanical Engineering, or Robotic Engineering departments

created to be a mobile system for persons with disabilities developed at the University of Delaware. In addition to the concept of semi-autonomy the system employs supervisory control using a planning system based on traditional AI techniques [63, 64]. The system integrates a user, simple visual processing of the environment, and STRIPS-based planner forming the interaction between the user and the robot. This may well be a solution for using systems such as STRIPS [43] on robots in the real world, but most autonomous robotics researchers have abandoned linear STRIPS-style planners decades ago because of simple task interaction problems (as exemplified by the Sussman anomaly). With such a planner the system is committed to having a user present to decompose plans into disjoint sub-tasks.

The field of teleoperated robotics has worked for decades on human-machine interfaces to enable an operator to control a robot remotely in space and in other distant or hazardous environments. There are a number of conferences and journals that disseminate this work [78, 90]. NASA initiatives have included virtual environments and semi-autonomy in that work [54, 35]. In the span of work in telerobotics, the sub-field known as "supervisory control", as defined by Thomas Sheridan [92] is most relevant to this thesis work. In the second edition of the journal Presence, Sheridan defines the term *supervisory* as it relates to robotics:

> Supervisory control does not require physical remoteness, though some degree of functional remoteness is inherent. In the strict sense, supervisory control requires that the machine being supervised has some degree of intelligence and is able to perform automatically once programmed.

This functional remoteness, and specifically the ability of the robot to perform some autonomous tasks, is what has been referred to here as semi-autonomy and it is that element that is sought in related work. However, the work in this thesis relaxes the notion of autonomy "once programmed." Here the programming is seen as an interactive task where autonomy is flexible and related to a specific context. Interactivity is considered a basic part of the system and in this way the system is seen to be collaborative.

The most related telerobotic work to this thesis is the subset of work employing 3D virtual environment technology for telerobotics. Though there are many similarities, few systems explore the same set of research notions addressed here. Investigators have worked on the interface between man and machine to enable an operator to control a robot remotely in space [78], and battlefield [10] applications and even used simulated environments for predictive task planning [65]. Because the body of research in the field of supervisory telerobotics and virtual environments is large, what is offered in the next section can only be a sampling of the work that has the greatest significance or influence to this thesis work.

Virtual Environments and Telerobotics Virtual Environments have been employed with telerobotics in a number of ways. By virtual environment, what is meant is a 3D representation of a robot and robot environment. Often, but not always, this environment is interactive and provides a means of probing different robot features. All work, by definition of telerobotics, is carried over some measure of distance or scale. Often the virtual environment is used to gain visualization of that remote space. Of the different systems that use virtual environments, the differences occur in the use, purpose and goals of the virtual environment vis-à-vis the system as a whole. A categorization of telerobot virtual environment use, with example references, is given here:

Workspace visualization: Allowing the human user to better visualize the robot's 3D workspace in a spatial manner. This might be a static representation or it might be with a form of real-time update of robot position [76].

Immersive telepresence: Virtual environments along with immersive display techniques (e.g. head-mounted displays) have been employed to enable a greater sense of 'presence', with the goal of offering the user a sense of being physically present in the remote space [119].

Programming: Using a model of a remote environment, the user can spatially program the robot by, for example, picking binding points for intended robot path or trajectory [65].

Workspace interaction: Allowing the user to interact with the robot and possibly setting up a model of the remote workspace. Sophisticated versions of this work allow the user to build and refine models of the environment [31, 29].

Rehearsal: Before executing a particular task, virtual environments can be used to perform a simulation and allow the user to possibly catch and correct any detected errors [34].

Most of the systems in the above categories, with the exception of attempts at immersive presence, use virtual environments in addition to other interface methods. If the interface consists of a graphical user interface, than the 3D virtual component will often be a window displayed alongside other robot controls. In contrast to this, one goal of this thesis work has been to create a system that is unified in the sense of using the virtual environment interface for the entire task. These different uses of virtual environments will be revisited throughout this discussion.

Visualization and interaction over scale In work with microrobots, the EPFL lab has developed an interface with an architecture employing a virtual

environment similar to the present system [79]. Here the robot has reflexes that allow it to respond to the local environment and uses the virtual environment as a medium for communication. Here the user can be an observer of the remote environment of the robot as well as an actor, within that environment depending on the mode of the robot. The purpose of that work is to allow a human supervisor to work at non-human scales, both micro as implemented, and macro as implied.

The work has been implemented on an arm-based Khepera micro-robot and the goal of the EPFL interface is to allow fine control via the specification of path-control points within the virtual environment. The authors also incorporate a model construction system based on a multi-modal system that incorporates range sensing and vision to build a model of the robot's working environment. Overall the authors use the model for specifying fine-grained control points for navigation and have yet to take advantage of the power of spatiality and deictic reference that the virtual environment affords for higher-level commands and greater robot autonomy.

Rehearsal The group at Sandia Labs has employed a virtual robot system which exemplifies the rehearsal style of employing virtual environments for robotics [34]. A system has been implemented containing a physical model of the environment and the robot dynamics so that rehearsals of robot tasks can be performed before running the robot on the real task in the physical world. The system is based around a gantry robot system that works in a large 20'x40' workspace. Users can be warned beforehand about impending motion problems. One of the more interesting motivations for this system is financial. The argument is that the virtual system will promote the sharing of expensive capital equipment. However in their presentation there is little attention or outside references given to the interface which is not atypical for reports on human user-robot systems.

One goal of telerobotics is the notion of *presence*, often attempted through techniques of immersion. The next section details what this might comprise.

2.4 Presence

Presence is its most basic definition is the subjective sense of "being there" or being *present* in a remote or virtual environment.

> The Astronaut would receive a sufficient quantity and quality of sensory feedback to feel present at the remote task site and would be able to carry out repairs or detailed inspections as if he were actually there [44].

Many studies have been performed that have attempted to explain the nature, origin and modification of presence in the environment and then also relate

these factors to real world practice [119, 45, 105]. This thesis views the role and degree of presence as a means to enable more engagement in the task work. This concept has also been a way of enabling *domain transparency:* seeing through the tool to the goal task [37, 56, 55].

Following Welch [119] the terms *telepresence* and *presence* will be treated as essentially the same. Welch *et al* hypothesize three factors that contribute to the sensation of presence. These are that the user:

1 feels immersed within the VE;

2 feels capable of moving about in it and manipulating its contents;

3 has an intense interest in the interactive task.

Most of these factors are the same as what contribute to engagement in the real world and continuing in that direction of transferring real world factors to the virtual the authors continue:

> The development of VE's [Virtual Environments] can be viewed as an attempt to produce by means of a computer program and accompanying hardware (e.g. a dataglove), the same experiences of clarity, completeness, vivacity, continuity, constancy, and presence that occur in normal perception.

This notion of continuity and constancy can be phrased as *habitability.* Through factors contributing to the sense of presence as outlined above, e.g. engagement, rationality, empowerment, an impression of the environment being familiar and "making sense" helps to enforce this sense of habitability. Also, in reverse, attempts to make an interface habitable should create a greater sense of presence. One approach to habitability is to transitively extend real world habits and competences into the virtual environment and work with the robot.

Heeter argues for three different types of presence: environmental, social, and individual [53]. Like Welch *et al.* these can be viewed as composing a tripartite space which represents the many variables that form the sensation of presence. Regardless of the structural description, these three components form an encompassing way to discuss the factors that contribute to presence:

Environmental factors: Range of sensory experience and modalities stimulated, amount of sensory resolution, degree of similarity between the observer's body and virtual representation, presence or absence of stereopsis, B&W vs. color presentation, presence or absence of perceptual constancy during movements, familiarity of the scene.

Social factors: Whether other (simulated) individuals are present in the VE, and the extent to which these others respond or interact with the primary observer

Individual factors: Assumptions that the observers bring to the VE, amount of practice they have had on the VE task, length of exposure and interaction with the VE, the degree to which they have become familiar with (and adapted to) the intersensory and sensorimotor discordances, individual predispositions to rely or attend to one sensory modality over another.

Clearly all three of these factors, environmental, social and individual, play a role in the sense of presence in a system including the one in this thesis. In particular, the isolatable environmental factors that are relevant to the sense of presence in this thesis work are related to the presentation and interaction with the graphical environment. These are that the physical presentation offers enough sensory resolution (e.g. pixels) and is sizable enough to display detail at a comfortable distance to the human operator. Other environmental factors include the number of modalities involved and what constitutes the interface.

In the thesis system the other individual present is the robot agent (or agents) and possibly other operators. There is then a social understanding built from the representation of the robot agent in so far as it displays a coherent view on the robot actions, capabilities, pose and scale within the environment. This is also where the beliefs and understanding of the user vis-a-vis the robot come into play. Does the human supervisor view the robot as an entity? A fellow co-worker? Is the operator working *with* the robot? Is the operator working *through* the robot?

Individual factors are by far the most difficult to measure. In order to gauge the influence of many individual factors, rigorous and well calibrated laboratory techniques have been employed to answer perceptual questions. While in-depth qualitative studies of individual responses may need to be performed to understand individual factors such as innate individual assumptions and predispositions. Despite these attempts, many of these factors still remain elusive and indescribable. There are quantitative studies that have attempted to address presence with some rigor (see [105] for an overview). Also individual factors will be, by definition, highly variable and dependent on each particular user. Most such quantitative studies concentrate on psycho-physical factors. Other factors such as cultural meanings, semiotics, social understandings might only be revealed by qualitative means. Since much interface work is based on the idea of metaphor with the real world, such individual and social cultural factors may be just as, if not more, important than psycho-physical factors.

One aspect that Heeter's presence factorization does lack is the identification of the interactive sense of continuity, rationality that may enable engagement and create a working interface. The factors to create these senses most likely cross the categorical boundaries in different combinations. For example, not included in the environmental factors is the breadth and type of modalities offered from interaction. These surely are significant for the sense of presence

and they are vital to the degree of control the user has and certainly involve environmental, social, individual factors. This is because the way we interact with machines is informed by the way we we interact in everyday social life (see section of Software Agents).

One more point that should be addressed in a discussion of presence is the effect of "lag-time." In many teleoperation systems, the reduction of lag, the time between a control instruction is given and when the effect is seen, has been given a great deal of attention. It has been supposed that reducing lag gives the operator a sense of 'being there,' and increasing the sense of *presence*. This is because the motive of many systems is to unify the operator and the remote robot as "one." In this thesis system, operator-robot unity has not been a goal. In fact, the human is seen as a supervisor and the robot as assistant, two distinct entities with different roles and responsibilities. The lag factors in the thesis system, though important, are not as relevant as is in many teleoperation systems. This is primarily because we have sought a higher level of autonomy for the robot than direct control. Lag time is important however, as it is in most computer interfaces, in confirming to the user that a command has in fact been "understood" or accepted by the system. In a deictic system this can be accomplished by the system responding with a simple "ok," "click," or "flash" when a verbal or gestural command has been accepted. The user will then understand that the command has been acknowledged even though it may take some time for the system to perform the request.

2.5 Virtual Environments

The field of *Virtual Reality* (VR) has been around for more than a decade.[5] In that time it has passed through the gamut of phases beginning with fascination, moving on to utopian promises and then to outright hype. By the time of this writing, much of the hype is subsiding, leaving the term 'Virtual Reality' a bit tainted, but yet enduring as a serious field of study in itself and still holding some of its fascination. Many of the most successful VR ideas have been adopted and incorporated into the research programs of other fields (*e.g.* Molecular Modeling, CAM/CAD, Telemedicine, Virtually Augmented realities, Teleoperated Robotics). In the field that is now virtual reality, some of the most promising work is in social applications. Examples of such social applications include graphical collaborative environments, virtual conferencing, and work applications such as shared CAD/CAM systems and room-sized conferencing and collaborative visualization. The work in this thesis is most influenced by these collaborative applications of VR, those are virtual worlds that can be shared and in which multiple users can interact. The particular type of system platform upon which the robot system is built is an instance of

[5] Although notions similar to Virtual Reality have been around before. Please see Myron Kreugar [67] for a view on pioneering work that precedes VR.

a Collaborative Virtual Environment (CVE). In this work we do not use head mounted displays and employ instead large screen display. The primary reason is to leave the human user unencumbered, while also providing a useful display and the possibility for collaboration between co-located participants.[6]

In this thesis an immersive virtual environment is used as the interaction medium with the remote world and the robot. Specifically the work is an application in the SICS Distributed Interactive Virtual Reality (DIVE) system [30, 50]. The Distributed Interactive Virtual Environment (DIVE) platform has been in existence since 1992 as a research software platform and has grown from a project based in shared multi-media communication. It is essentially a distributed and shared virtual environment system based on multicast and a peer-peer model for sharing and updating the environment model. It has a high capacity distribution infrastructure to support multiple users and supports a number of common network standards at different levels (*e.g.* TCP/IP, ATM, URL, MIME) as well as supports a number of file formats (*e.g.* VRML, VR, AC3D). Application functionalities can be programmed in C, OZ, Tcl and integrated with other tools, (*e.g.* Netscape, Politeam for document sharing, VAP for audio conferencing).

Recent research in immersive virtual environments and human-computer interaction at SICS has worked on building a framework for "natural" interaction. Aspects of this work are the study of interaction between agents, human and others *in* a shared virtual environment [15], or the construction of mechanisms for human users to interact with the virtual environment [60]. Some of the work in this thesis, the 3D mouse pointer in particular, has come out this work with "natural" interaction [51]. That work has tried to elicit possible meanings for natural while realizing that there may be no one definition of "natural." The definition depends on context and varies in respect to such factors as the culture, the task, for the individual's abilities. Instead a concept that is far more complex and less generalizable may be more suitable, such as the concept of "appropriate." Thus the existence for a general interaction framework comes into question and leaving the need to create interfaces specific to particular tasks and contexts. This idea of appropriate interaction does share some resonance with modern trends in computer infrastructures [117] and interfaces [82]. One simple approach is to search for multiple methods of performing a task, each context dependent.

2.5.1 Enhancing environment visualization: Augmented Virtuality and Augmented Reality

Augmented Reality (AR) is one branch of research in computer generated immersive environments that has found a number of promising applications. The

[6] Co-located collaboration is not directly addressed in this thesis, but the concept has guided some of the interaction decisions and a new project co-written by the thesis author explores computer-supported 'shoulder-to-shoulder' collaboration.

visualization sub-system model presented in this thesis (*Reality Portals*) has the capability to use a base 3D polygonal model of the world and with the aid of calibration overlay video into the graphical world. It also includes the possibility with the slightly modified routines to do the opposite, to enhance a video image with correctly position graphics. With such systems it is possible, for instance, to enhance or augment the video stream with graphical information that may not be visible but may be useful. This overlaying and mixing of graphics into a video stream is referred to as Augmented Reality (AR). The complimentary operation of laying video onto graphic worlds has been referred to as Augmented Virtuality (AV).

The classic examples of Augmented Reality employ a display that is worn on the head, that enables the bearer to see both the real world and a graphics display that is overlayed onto semi-transparent screen. In this manner a user of such a system would see the physical world augmented with "appropriate" graphics. A user might use such a system to repair a laser printer [39], or repairing an automobile engine with annotated instructions appearing on the lens of the see-through glasses [66]. In the previous example the user was present in the physical environment. Similar operations can be performed remotely. Milgram *et al* have used a telerobot system delivering video of a remote scene to a special display worn or observed by the operator that is enhanced with interactive graphics. Such applications include virtual tape-measurement on real scenes as well as graphical vehicle guidance [75], and enhanced displays for teleoperated control [73]. In addition to this standard notion of AR a virtual environment can be embellished with real-world images.

Milgram [75] has attempted to map out a taxonomy of "mixed realities", those that blend graphics and real world video. One axis of this taxonomy spans from Augmented Virtuality to Augmented Reality, with citations and positioning of much of the work in between. In addition, Benford *et al.* have created a similar mapping of these techniques and applied them to the field of entertainment and in particular to employing CVEs for inhabited and interactive television [18]. A place where television streams, video conferencing, and traditional collaborative virtual environment systems coincide. The base techniques for all these systems are the same and include models of the real and virtual scenes, methods for locating views (camera and graphical) within those models, and methods for mixing video and graphical streams.

Corby and Nafis describe an ROV that roams a nuclear core for the purpose of inspection [59]. They have motivated the use of a robot by stating that it can be employed without causing reactor downtime. Also human core inspection is problematic and would be considered a *dangerous* environment. The cost of nuclear reactor downtime is enormous (*e.g.* when measured by the metrics of money and quality of service) and must be minimized. To perform the task, their system includes a three scene display, with three points of view including the position of the robot within a map of the core, an on-board robot video camera, and a synthetic 3D graphical view. In their work, they cite having

significant problems with having video as the only source of visualizing the actual remote environment. It is that reason that motivates their three-screen solution. The following quote from their report highlights the problem which elements of augmented virtuality, such as as Reality Portals seek to overcome.

> The field of view is often very small compared to the size of the reactor. [...] The path that the ROV must take is often a very tortuous one involving many translations and rotations to enter restrictive areas. It is often impossible to determine the next step based only on the video image coming back from the ROV camera. Finally, because the camera is mounted on a 2 DOF arm, it even more difficult for the inspector to verify that the current image is of correct area viewed in the expected configuration. *N. Corby & C. Nafis, GE Corp R&D [59]*

This is to say that merely having the video of a remote environment does not afford an understanding of the physical structure of the remote environment. One approach to a solution is a combination of graphical and video as in AR and AV. One specific solution is to enhance a 3D virtual environment, with appropriately positioned video textures providing an environment with the greater physical structure of the remote location and the details from the video images. Reality Portals are an instance of such a system.

2.5.2 3DTV and virtuality

The Reality Portals AV work can also be viewed as an instantiation of work in the general area of immersive 3D video, or 3DTV. In this immersive video, a user is not restricted to the 2D plane for video but can interactively choose views and explore video regions. For our purposes, we see the Reality Portal system as a means of filtering out non-essential details from a potentially cluttered real world scene. Institutionally supported programs in the field of 3D video are being carried out at UCSD by Ramesh Jain's group [61], and also at CMU by by Takeo Kanade's group [88]. Both of those applications have focused on creating mixed realities that can be interactively accessed. In this work we have created a tool that can be used for near real-time remote investigation. It is "near real-time" because there are some delays introduced by the segmentation and texture application process. These delays are approximately 1-3 seconds in the demonstration prototype. Though there is significant room for optimization in the prototype. Kanade's research is intended for broadcast applications and Jain's lab has started to explore real-time security applications partly as a result of communication of the work on Reality Portals by the author.

2.6 Software Agents and Robot Assistants

The research and development of software agents has grown in the past decade from a sub-field inside AI to a bona fide area of its own with workshops, symposia, proceedings, books, collections, research prototypes and commercial applications. There are elements in this thesis work that are related to some of the work in the field of software agents. The relationship centers around the design of an embodied assistant. These similarities are more clearly seen from the software agent researcher's point of view than from the robotics researcher's. There are software agent researchers that see very little difference between their work, and that in intelligent robotics [112]. This section, rather than being a complete survey of agent related work, presents how the field of software agents is related to the present thesis work, a number of factors that go into the design of agents, then presents a contradictory argument to contest that view that work with software agents is the same as work with physical robots, and then steps back from the differences to point out the implications.

2.6.1 Derivative software agents

One reason for a connection between the work in the *software agent* (SA) community and that in the robot community is that key persons working with software agents have come from the field of robotics and AI (*e.g.* Pattie Maes, Walter Van Der Velde, Barbara Hayes-Roth). Thus some of the the work in software agents has built on work with artificially intelligent robotic agents. This includes architecture models for software agents that have come directly from the models developed within the field of artificial intelligence [37, 112]. In fact, as evidenced by the following quotation, many researchers in the agent community frame their work with respect to robot work.

> The idea of an agent originated with John McCarthy in the mid-1950s, and the term was coined by Oliver Selfridge a few years later, when they were both at the Massachusetts Institute of Technology. They had in view a system that, when given a goal, could carry out the details of the appropriate computer operations and could ask for and receive advice, offered in human terms, when it was stuck. An agent would be a 'soft robot' living and doing its business with the computer's world. *Alan Kay* [62].

For the purposes of this section, this quote by Alan Kay is offered as a definition of software agents. Also in this quote, the connection of an agent being a "soft-robot" is made explicit. In the literature already cited, there are frequent comparisons between robots and agents. The above quote also points out the human-interaction aspect of work with software agents. It is this aspect that most separates the work in software agents from that in artificial intelligence. This difference aligns the software agent community more with

the human computer interaction (HCI) community than it does with the AI community. Identifying this split brings up at least two comments. First that at least part of the software agent community is derivative of AI community and second, many of the hard questions about AI techniques may have become irrelevant in the constructed environments of user-interactive software agents.

It is in the aspect of a computer-based system interacting *with* a human that ties this thesis work with the concept of a software agent. In both systems the agent is in service of the human user, it is the human that provides the primary goal, and there is an interaction discourse between the human user and the agent. In particular the supervisor-assistant metaphor of teleoperated control is an example of such a human-agent interaction system.

However, claims that interaction with robots can be simplified to interaction with software agents (and later perhaps generalized to the hardware robot) are mistaken. The primary differences can be seen from various perspectives. One difference lies in the root of the debate of new AI vs. GOFAI (Good Old Fashioned AI), e.g. the behavior-based school and the goal-oriented school. This debate has also been framed as a top-down vs. bottom-up division of approaches. In that debate it is argued that robots need to be embodied from the beginning, the solutions developed in simulations will not transfer to real robots [27, 87]. In this thesis, the term *agent* is mostly not used outside of this section, and instead 'robot assistant' is employed.

Interactive software agents have triggered a large debate that has yet to occur in robotics. Though it may come with greater domestic deployment. Some of the issues in this debate are just as pertinent to the design of interactive robots.

2.6.2 Interface agent debate

Human Computer Interaction researchers debate whether intelligence at the interface is good HCI design [96, 56, 72]. In terms of teleoperated human-robot interaction it is "intelligence" (or autonomous competence) that will distinguish autonomous or semi-autonomous teleoperated robots (supervisor control) from a purely remote-controlled application (teleoperation). The controversy centers on where the designers place the "intelligence." In this thesis work it is not intelligence in the interface, it is intelligence (however limited), in the form of autonomy, in the robot that the user interacts with. The interface, as a system, is a way of accomplishing a set of tasks.

There is a debate about software agents that is often heated. Lanier writes "The idea of intelligent agents is both wrong and evil [...] I believe that this is an issue of real consequence to the near term future of culture and society" [68]. Much of Lanier's argument against agents centers on the fear that users will be "dumbed down" to the level of the implemented agent and that the user will somehow become more driven by the whims of the forces that created them. These are not simply concerns of the polemic, those on the academic side of

HCI are voicing related views about agents. One example is Ben Shneiderman when he writes:

> I am concerned that if designers are successful in convincing the users that computers are intelligent, then the users will have a reduced sense of responsibility for failures. The tendency to blame the machine is already widespread and I think we will be on dangerous ground if we encourage this trend. [96].

This is similar to what Foner refers to as the *social contract:* "Most [commercial agent offerings] tend to excessively anthropormorphize the software, and then conclude that it must be an agent because of that very anthropomorphization, while simultaneously failing to provide any sort of discourse or 'social contract' between the user and the agent" [46]. It is this discourse that is both hard to define and hard to design. Phoebe Sengers has addressed part of this issue by pointing out the lack of attention to the signification of agent behaviors from the *user's perspective.* She writes: "what matters is not the internally-defined code as understood by the designer but the impression the agent makes on the user [91]." The system constructed by Sengers gives special attention to the subtle cultural aspects of agent behavior representation within the context of use.

One aspect emerging with the research and development in software agents is that there is more to the interface than efficiency, that there may even be entertainment value. However these again do not have to be polar opposites work vs. recreation, there are notions of 'fun work,' and it may be those experiences that are quite desirable. This may in fact contradict some of Shneiderman's mantras about direct manipulation and "predictability" "controllability" and "mastery" being the primary element of concern in an interface [96] and point to other less quantifiable interface properties. This is to say that there is more to the interface than efficiency and "time to completion" may not be the best way to measure the success of the interface as a whole [56]. There are other qualitative aspects that need to be taken into account. Such qualities are much harder to identify and have more to do with personalized notions of productivity, intuition, and consistency.

2.6.3 Attributes vs. Endowment

This debate about the "good and evil" is as much about what users understand the interface to be as its actual constitution. This leads to a further discussion about user endowment and system attributes. There is a delicate balance between appealing to a "life-like" metaphor in a software agent and, while in the pursuit of that goal, promising too much competence in the presentation. As an example of the power of this seemingly subtle difference, the debate about whether or not the graphical agent for the Olga project had 'antennae' or not posed a severe threat to project cooperation. The reason for adding antennae

was to add a visual cue that the agent was not a human, the opposing view was to purposely make it look more human [110]. What the user puts *onto* the system is endowment, what the system actually has capacity for is its attributes.

The warnings from Shneiderman and Lanier mentioned above are partly centered on the idea of personification. Presentation may lead users to *endow* the interface with other traits *e.g.* if the representation is human, there may be an expectation of other human abilities. Inattentive presentation may lead to false expectations, leading to disappointment, leading to rejection of the interface and tools. A related warning is that the presentation of autonomy in the interface will in the end "disempower the user." This may result from a sense of intimidation from competence the user endows the interface with. These warnings point to the power of the interface. The Eliza-syndrome and the Turing Test are demonstrations of this power. The intended effect of providing such an interface is to achieve a form of cohesion or habitability in the interface and in the dialogue that is effected. Thus a designer may not be able to simply refer to control and mastery as goals, many of the desirable properties of an interface may be subjective.

This distinction between endowed intelligence and innate intelligence is an issue that arises quite often in AI and in robotics as well as in the field of software agents. This is precisely what Horswill built into his Polly tour guiding robot system [57]. Upon start-up, from the robot's speaker came "Hi, I am Polly I am the MIT AI lab tour guide robot! ... and I do not understand what I am saying right now." The robot would then move around the space and conduct a tour. With this statement it was made clear that the robot did not understand its own speech, and also, by inference, made it clear that it would most likely not understand a user's speech either. What might appear to be a touch of whimsy from the interface designer is actually a very strong interface statement about the robot's capabilities, helping to define the assumptions under which the robot operates its tours.

Another example from the same lab is the 'Cog' robot project. Regarding a session of taping of a demonstration video, Rod Brooks had the following to say:

> "Cynthia and the robot were taking turns. The robot had no notion of taking turns, but Cynthia knew how to take turns and she capitalized on the dynamics of the robot. [..] It seems to me that people cannot but help themselves interacting with artifacts the way they interact with people."[7]

Rod Brooks indicated that the turn-taking seemed to be an activity that users endow on a robot. Thus the fact that the human and robot take turns is

[7] Quote is spoken by MIT AI Lab professor Rod Brooks while describing his *Cog* robot in the 1997 Discovery Channel documentary, *Robots Rising.*

an emergent property of the system. In this way the tendency is for the human to socialize the robot. Many more examples of this phenomena can be found in the Software Agent, HCI and AI literature [114, 56, 74, 37]. The point is to make the difference between attribution and attributes. It should be clear that much of the intelligence that is endowed on the robot comes from the user and to some extent, this is part of the goal. However, the limitations of the system also need to be represented. This idea of showing the limitations has been described as a form of transparency. Two categories of interface transparency are *domain transparency* and *internal transparency*. Domain transparency is being able to see through the tool to the task. Internal transparency is to be able to see the workings of the tool. A further discussion of these concepts as regards the construction of intelligent agent interfaces can be found in [56].

Anthropomorphism is not the only design dimension to trigger attribution on the part of the user. There may be far more fundamental social responses at work. These social responses may be present beneath the conscious level. Stanford researchers, Reeves and Nass, have studied human social behavior with inanimate objects. Some of their results claim that whether people want to or not, they will treat non-animate objects, (*e.g.* computers), in a similar way to how they treat other humans. Thus personification of the interface is not necessarily the most critical factor. It is more the discourse that is taken between the human and the computer. One example of this is politeness. If a computer is polite to a user, it is quite likely that the user will be polite in return, even when that behavior may be inappropriate. Users will apply these responses unless there are strong indications that that behavior is inappropriate. Thus the complementary concepts endowment and attributes may be a fundamental to human nature, a natural social response. Reeves and Nass also believe that people cannot be introspective about their interaction with computers. In fact, users will often deny their social interaction with computers [89].

Many of these points about the interface, coherence, attributes, endowment, transparency and social interaction with respect to the system in this thesis are taken up again in the Chapters 5 and 6 where the a study of use is presented.

2.6.4 Differentiating robots and software agents

There is an dichotomy regarding the relationship between mobile robots and Software agents. This dichotomy exists between the academic positions of those in robotic research and those in software agents and how they view what an agent is. It is not difficult to find literature that relates software agents to the research in robotics [112, 37], but the converse is difficult. In fact, some in robotics have denied the relationship to the point of it not being important enough to have a position on. From the software agent community the view that software agents are simply "soft robots" or softbots, is held even from those that are aware of the distinctions of old AI and reactive robot systems.

A primary distinction of the latter community is the insistence that robot applications must run and be tested in the real world.

One perspective on the differences is ecological. Even though an agent may be connected to a network accepting unexpected information as input, that agent exists within a well defined computational world. Its situation is different than a physical robot in the real world. The agent's universe of operation is controlled, more predictable, and in essence, modelable. Given a particular set of inputs to the agent, e.g. from the human user, it should be possible to develop a computational model of the agent's input-response cycle. With a physical robot, operating in a real and possibly hostile environment, this computational model is far less tangible. This is paradoxically complicated by the fact that robots are built to be robust.

Robot inputs are not easily modeled and further that problem may prove to be intractable. Very often sensors are noisy and imperfect and the science of understanding sensor data from the real world is far from complete. Sensor devices themselves can be faulty or affected in ways that are not detectable by the software that runs them. Thus input to the robot may be undetecably transformed. In addition, the real world is a complex unpredictable space, it is unlikely that any predictive model of real world composition and events will ever be developed. Robot output, e.g. actions, may be hampered by environment unpredicted features and events, or by undetectable failures in the robot system. Internally it may be possible to build a robot that has very good diagnostics, but it is likely to be impossible to detect every intermittent sensor fault.

Simply put, there are differences between software agents and robots, and the software to run one is not necessarily transferable to the other. Most of this is due to robots working in often harsh and unpredictable environments. Assuming the opposite, that they are the same, just in different forms, is making the same, likely false, assumptions that members of the GOFAI (Good old fashioned artificial intelligence) community made for years. This is, in essence, the basic Cartesian separation of mind and body. Few in AI research have taken a critical perspective on this process. One of the few exceptions is Phil Agre's well presented argument that a Cartesian-inspired cognitivism has severely hampered AI research [4]. To some extent the new AI and behavior-based robotics has broke the separation of the mind and body and moved toward an interactionalist and ecological approach.

The cruel real world The difference between robots and software agents lies primarily in their computational interface to the world. In robotics this is most apparent in research in the reactive robot community. In this community there is an insistence that robot research must be carried out and tested on real robots that run in real environments in the real world. This community shares a philosophy with the Active vision community that insists that computer vision

algorithms must be run on real cameras in real settings with active pan-tilt heads, instead of in simulated environments or static images [11, 115]. The way a reactive robot interacts with the environment is through its sensors and its actuators. The methods by which this data is gathered, the way the models of the world or models of actions are built up, are fundamentally different than those of software agents.

In most research on software agents, the interface to its working environment, or domain, is well known. Recognized inputs can be mapped out, actions relating to those inputs can be specified, and a number of outputs, or actions, can be taken. The view of an agent then is that of a process operating in a self-contained environment that can be well described. Contrast this to robotics where one of the main foci of research is the field of recognition of its environment, a first step to generating a set of inputs for decisions on appropriate actions.

One can perform a thought exercise to compare the work of the software agent and that of the mobile robot from the perspective of the research developing the system. A software agent system might consist of a standard desktop system: a monitor, CPU with storage, and keyboard and mouse interface. Displayed on the monitor is a representation of a graphical agent. This graphical agent is the user interface of a program that is attempting to help the user at some task based on the user's task history. This is not an uncommon configuration for a software agent. Then imagine an analogous robot configuration, consisting of a mobile robot, with sonar sensors, an arm, and an environment that is a mock-up living room, designed to perform some fetch task. To demonstrate the statement that key differences between mobile robots and software agents lie in their interface to the world, imagine what happens when unexpected inputs occur due to a problem in the input system.

For the software agent, let us imagine that a key part of the system is a trace of the mouse over parts of the monitor screen. The user employs the mouse to focus on different screen elements. The software agent in turn uses this focus history to develop its model of the user. Imagine that the particular surface that the mouse rests on is slippery. In this particular case, there is enough error in mouse tracking to inhibit the movement of the pointer and cause the agent to collect false interest foci as the user errantly moves the mouse on the screen to the location she wants.

Now imagine that an analogous hardware problem develops in the robot so that the wheels on the robot slip on the surface, causing encoder positional reports to be in error. The cumulative effect is so significant as to cause the navigational system problems. It is quite possible that such a navigational difficulty would be so significant as to force the designer to redesign the navigational algorithms or to consider modification of the floor so slippage does not occur.

The question here is how the robot and software agent researchers react to their specific problems. Does the software-agent researcher consider this an

'unfair' test of the system? Does he assume that problems like this should not exist and that they can always fix such problems (e.g. with a mouse-pad) and start again? With the robot researcher, the above situation will most certainly be treated as a trial for robustness of the system. The key difference is that the interaction between the system and the world is a significant element of robot research. This difference in how the researchers respond is not superficial, it actually goes to the core of what reactive robot research is and how it differs from some traditional robot and AI research. One reason this is not a trivial distinction is that *embodiment,* situating the robot in the real world, has become a key distinction between styles of building mobile robots. A number of researchers have discussed this distinction between styles of robotic researcher and its implications much further [108, 106, 87].

This argument only wishes to point out that robots need to work in the world and it is impossible to re-engineer every location to suit the robot.

Implications One difference then between this thesis work and software agent work is that, while the agent community is trying to create an interface that is personalized, possibly identified as an entity with human qualities, here the attempt is to humanize an interface and create channels of communication. Such an approach can create an interface that capitalizes on the capabilities and practices that already exist for human shared work practices. One lesson is that designers should consider the disadvantages of endowment when designing systems purported to be autonomous.

2.7 Computer Supported Collaborative Work

Although human-computer interfaces have existed since the machine was invented, it is only in the last few decades that the field of Human Computer Interaction (HCI) has matured and separated into its own discipline. This has included a call for greater awareness of the need for attention to design and detail, human-centered interface and interaction design has been noted by many [83, 95, 81]. Not every concerned discipline has heard this call, much HCI work still remains removed from disciplines that could be informed by it, *e.g.* robotics.

For this thesis work with robots it is primarily the work in the CHI sub-field of Computer Supported Collaborative Work (CSCW) that has most inspired the research. CSCW research examines the way computers might support traditional and new forms of collaboration. Within CSCW there are is a culture of examining actual practice and applying the implications of those studies as to CSCW systems [25]. In addition there is an emphasis on cooperative systems as seen in context of use [109] and a tradition of working with users in the design process [23].

An emerging sub-field of CSCW is that of Collaborative Virtual Environments (CVE). In the last few years this community has emerged to have its own ACM sponsored biannual conference (*e.g.* CVE2000). It is in this field that the thesis author is most engaged and concurrent work to this robot work has explored 3D collaborative desktop representations [38], and supporting co-located collaboration for children [14]. It is that CVE research context which the author brings to collaborative robot systems. CVEs brings a number of issues to bear on how persons collaborate and how computer systems might support this collaboration within the support of some form of virtual environment, graphical or overlayed on the real world. Much of this work has traditionally been informed by studies.

The tradition of performing ethnographical studies in CSCW and CVE research attempts to expose the complex ways that that individuals weave together both individual and collaborative activity in real work settings and then asks question about how these activities might be supported with computing technologies. Collaborative work in real settings often involves subtle cues. For example individuals often wait for signals that another is finished with one task (*e.g.* and exaggerated push of the return key, pushing the calculator aside, etc.) before initiating a new collaborative task with that person. Another feature is "outlouds," statements made to no one in particular, but that provide background information that anyone might respond to, and to which the caller does not necessarily expect a response. Another feature is that collaboration does not always require explicit agreement but might be brought about through a *stepwise progression to collaborative activity* that may not be explicitly understood as such or the same each time [52].

Collaborative work often depends on concepts such as tacit knowledge and what has been referred to as a "working division of labor," where: "the division of labour is organized dynamically according to need and does not necessarily follow a prescribed form [58]." It is often these subtleties that keep workers "geared into" what they are doing, keep each other in synch, while also maintaining a high engagement level and a need to make work sharing explicit can detract from the level of engagement. For some applications there has been a fear of over-automating processes for fear that automation may remove some of the engagement and awareness of the state of things. For example, in a recommendation for an air traffic control system design controllers were forced to manually sort the progress strips used in maintaining an awareness of the air traffic, though this action, controllers were forced to be engaged in the activity [19].

This work can in fact inform a system designed to enable a human to perform a task with a robot. There is a need to consider enabling the sort of subtle features that can help maintain high engagement and lower frustration in communication of the task (as opposed to performance of the task). This is to say that functionality to enable collaboration can be made implicit and only require tacit agreement. It is through those features of the system that

an attempt is made to keep the human in focus on the task and not on the interface or the robot.

Hollnagel makes a now classic point about working *through* (instead of *with*) interfaces to the task [55]. This has been referred to as *domain transparency* and is meant to convey the idea that the tool does not take focus away from the task. Sheridan has made the point that the user works through the robot to the task [92]. However when the robot has more autonomy than in a teleoperation situation this comes into question. Should the human work *with* the robot through the interface? This distinction depends on the task and the competence of the robot and perhaps the distinction is not made explicitly, but offered through a number of tools so this sharing is flexible.

Division of labor One might conceive of a clean border between work performed by the robot and the human. However, boundaries that define the divisions of labor are more fluid in nature and can shift. Some systems have been built which can be seen to support this concept of task sharing. Though they do not make these arguments about collaboration, Burtnyk and Greenspan have introduce the concept of *interactive modeling* [29]. This enables a user in an augmented space to "develop and refine a quantitative, but necessarily incomplete, model of the portions of the remote world." Here a notion of interaction allows a human and a robot to construct a working model of the robot environment. The model is working in that it is in a state of refinement or greater change depending on events and task goals.

Another model of the division of tasks are further addressed in the work Milgram at the University of Toronto [76]. In their report they point out:

> To overcome such problems [encountered in unstructured environments], realistic teleoperation tasks should be appropriately allocated between the human operator and the machine intelligence, such that the respective capabilities of each are efficiently utilized. The reasoning behind such a division of responsibilities is that in general, humans are more suited for higher level perception and reasoning, task conceptualization, understanding the environment and dealing with unusual circumstances, while machines are good at low level sensory control functions, precision, reliability, and computationally intensive tasks. [76]

The above statement indicates a similar direction and user approach to that which motivates the system in this thesis work. Another approach to this sharing of task work from the University of Bielefeld. In their project to build a *Situated Artificial Communicator,* researchers have combined a two-armed robot, task planner and vision system to construct a system to build Baufix models with a human. In this system the human offers the high-level guidance and is available for "cooperative plan execution." One of the main purposes

in involving the user in this system is to help resolve robotic failures. The result is a system that has a tight coupling of the planning module and the dialogue system so that the human can intervene when things go wrong as well as provide the next steps in construction. This is a sophisticated system and involves a great deal of attention to detection, identification, and interactive solving of robotic fault during cooperative construction. The result is a system that attempts to provide a natural form of dialogue through speech. This work shares a philosophy with this system of involving the user with a robot in a cooperative task[48].

For the next generation of autonomous remote vehicles on Mars, NASA has been exploring the concept of "adjustable autonomy" [35]. The goal for design is to enable users to "interact with the systems at whatever level of control is most appropriate whenever they choose, but minimize the necessity for such interaction." They refer to the perspective as *human-centered* meaning that the first principles of the system are the interaction and the framework for system design is built around that. From that point it is safety, effectiveness, and cooperation with humans that are emphasized. The framework includes planning for traded and iterative shared planing and the planner design is intended to make the human aware of tasks where the human might offer assistance. Their framework requires a model of human competence to "know" when the robot needs assistance. This project is at early stages, and no user study reporting has been yet seen (though surely it is on the way). Like the flight traffic control system, there needs to be attention to the subtle features in a system that maintain engagement on the part of the human. Full autonomy may not be the goal when interacting with expert users (chemists, engineers, geologists, etc.) back on the Earth's surface.

Many systems tend to make collaboration an artificially explicit activity that is given by making specific commands or menu choices. When designing collaborative interactive systems it is difficult to be informed by ethnographical studies. One representation of this difficulty is has been that of a distinction between the fields of sociology and systems design, one is analytic and the other synthetic.

> "The fundamental approach of each discipline is totally different. Sociology is analytic. It is concerned with gathering and interpreting data about some social situation or process and drawing some conclusions from that interpretations. By contrast, software engineering is concerned with synthesis- designing and building new abstract models of the real-world. Thus, sociology focuses on and pays great attention to detail; software engineering strives to hide detail through abstractions. " [58]

The position of this thesis is that it is important to be aware of these concerns, and one of the best methods of insuring this is by performing studies of prototypes together with users. It is through these studies that features, such

as those from background ethnographic studies of practice, can be identified and strengthened or weakened as appropriate. This topic *vis-à-vis* specific design concerns is brought up again in the Study of Use and Findings chapters.

2.8 Robot User Interfaces

It can be argued that HCI has recently only begun to be taken seriously and it will take radical changes in the way technology is developed for most computer applications to become truly "human-oriented [82]." Though this is true for computers in general, this is especially true within the field of robotics. 'Intelligent' robots, those with capabilities beyond programmed precise factory floor routines, are not yet mature in any market. The first versions of autonomous lawnmowers and vacuum cleaners have only begun to appear and issues of their acceptance and interaction are at very early stages. As the technology and ability to provide service robots matures, there will be a need for HCI principles in the design process. To date the work in human-robot interaction has been mostly dominated by technologists, *e.g. ROMAN* and *SPIE Telemanipulator and telepresence technologies* conferences, and the journal *Man, Machine, cybernetics*. With the exception of the journal *Presence* the level of interdisciplinary work found at conferences such as ACM CHI and CSCW/ECSCW has not yet found its way into the robot community.

Usability and robotics do not traditionally go hand in hand. Even as usability and user studies are becoming more common and the robot-human interface must be seen as part of the system as a whole, usability experiments are the exception rather than the rule. This section examines four human-robot systems that have a focus on a specific features of the human-robot interface issues in this thesis. These systems are:

Model-based supervisory control Blackmon's system involves a 3D virtual environment interface to a robot arm. Studies have been performed and some of the ideas of this system have been used by NASA Ames in visualization of the Mars Rover.

Teleassistance: Deictic Robot Control Pook's system employs speech and pointing gestures for control of a robot arm.

MissionLab Toolset The MCS from Georgia tech is used to provide supervisory control over a number of robots. A user study has been performed to examine some efficiency issues of the system.

Multi-Agent Supervisory Control Adams' MASC system is a human interface for control of multiple robots. A user study has been performed to evaluate efficiency issues.

While a number of systems have features in common with the system in this thesis, few are known that are complete systems. This section focuses on these four specific research systems that do attempt to make the human interface a core part of the robot system and are seen to have similar goals to the work in this thesis. Each shows a different element of commonality with the system in this thesis and each of these systems and aspects are discussed in turn.

2.8.1 Model-based supervisory control

Blackmon has built a system for supervisory control based on a virtual environment [22]. The main application for the system is the interactive planning and rehearsal of tasks prior to issuing a command to the robot. While the robot executes the task, the supervisor can then take a monitoring role. The robot involved is one, or multiple, fixed robot arms within a structured task environment. Most of the work has been at a desktop workstation with some special interaction devices such as a joystick. In their system, provision is made for selection between different control methods, such as direct remote manipulation and supervisory control. They have also addressed the importance of usability testing and the role of presence in the interface.

In their user tests of the system comparisons were made of different methods of displaying the model as well as different methods for control. In a study of display techniques they tried to gauge the effect of realism of the display with the sense of presence it conveys. They found that the increased pictorial realism increased the sense of presence, and delay of feedback diminished it. Also constant activity increased presence, even if the feedback or pictorial realism was diminished [119]. Pictorial realism in that system was the degree of verisimilitude of the graphical image, not a video enhanced graphic image. In the comparison of control modes, manual versus model-based control were compared. In that study they found that although the learning rate was the same for the system with the virtual environment versus the purely manual control, the model-based approach reduced collision errors and supervisory control had the shortest time-to-completion [21].

Their work with a virtual model-based approach also sought to validate and build on a model of visual attention, the scanpath theory of gaze [107]:

> Decades of previous research have shown that model-based approaches can certainly assist the human operator in teleoperation tasks. Visual aids in the form of a separate graphic model with arbitrary viewpoint or as a visual enhancements superimposed on the transmitted video images facilitate spatial perception of the remote site.

Visualizing the robot environment Originally the model-based system used only peripheral camera views, and did not integrate the camera views into

the virtual model. In their laboratory environment visualization work, stereo projection and two robots have been employed. In comparison the system in this thesis work inserts video directly into the virtual reality model through the technique of Reality Portals. Blackmon developed this direction in later work on the Pathfinder project. During the Pathfinder Mars Rover mission VRML models embedded with textures were used to visualize the remote environment. In the Mars Rover work, Blackmon took the opportunity to integrate some of his visualization techniques with the data received from the Mars Rover. What was created was MarsMap a graphical VRML model of the Mars environment augmented with camera images form the rover [20]. This was made available on the WWW and showed one of the first generally popular applications of virtual environments and robotics. Both that work and the work in this thesis have been influenced by cross-discussions and presentations of the methods.

In their work with the virtual environment, Blackmon has used traditional GUI panels and menus. In this thesis work, there has been a deliberate attempt to unify these into the 3D environment. It is the belief, gained through informal studies, that 2D windows floating over the 3D display are distracting and erode the sense of presence.

The model-based supervisory control work overall is quite similar and influenced this thesis work. They have taken positive steps in looking for evaluation of different methods of display and visualization and have shown results that indicate that a virtual environment can be used to create an easily understood interface that might be more effective under certain conditions.

2.8.2 Teleassistance

Pook describes a system developed to demonstrate the idea of *teleassistance* as distinguished from other forms of human robot control. Teleassistance is a real-time control strategy where the robot is shown the relevant objects of the task and given hand gesture commands to perform on those objects. In this way the system is deictic. Pook's thesis introduces a version of qualitative control, a strategy that is designed to be flexible and to allow human interaction and direction. This system is applied to the application of flipping an egg in a frying pan. Pook argues that this type of system solves a number of problems found in the technique of "learning by showing" and even makes some brief arguments on why this type of system is a necessary alternative fully autonomous or fully controlled teleoperated robot architectures.

Although the work in this thesis is sympathetic to Pook's stance, it is a difficult position to rigorously present and defend and such an extreme *user must be in the loop* position is not taken here. However a number of problems with other forms of systems can be shown to disappear by introducing the human operator at a certain level. Taken in that context, those problems would be issues of delay, saliency, current state of robot autonomy. Teleassistance does seek to answer some of those questions within the framework of a robotic hand

manipulation system controlled by a human operator using a VPL dataglove and tracking system to indicate gestures to the robotic system.

In the presentation are models of the human nervous system and their relation to particular strategies for human-robot control. Pure teleoperation is compared to the concept of the *homunculus*[8]. The homunculus view, as a control strategy is the view that there exists a 'little man' inside the brain that controls the body. Pook argues:

> "The two engineering flaws of the homunculus recur in teleoperation. 1) the teleoperator bears the computational load, controlling every movement and noting all feedback. This is exhaustingly tedious for the operator and highly susceptible to error, such as inaccurate or inadequate feedback, poor motion mappings from human to robot or operator inattention. [...] Remote control causes even more problems for the teleoperator than for the homunculus. The homunculus at least cohabits the body of its robot; the teleoperation is off to the side in a different body, perhaps even at a remote site. Not only does this cause communication lags, but it can skew visual perspective, limit the feedback available to the operator, and make mappings awkward and incomplete.

Pook also raises the question: "Human supervisory systems have long been in use (e.g [41]) but the question arises of where to draw the line between human cognitive control and robot servo control. In other words, where is the conceptual interface between human and robot?" This is related to the questions about division of labor.

Pook answers this by dividing some of the literature up into what she terms "pragmatic" and "learning by watching." The first is expressed in systems that use menus to select actions and or constrain the motion of the robot to certain geometric primitives, even if the operator deviates from these paths. The second are those methods of "showing" the robot the task that needs to be accomplished. In common industry practice the later is done by using a "teach pendant" to first define the key points of a task, e.g. insertion, removal, placing locations, and then refining the commands until the robot performs the task acceptably. The robot is then configured to repeat this task. Each run is thus the same up to the limits of the robots calibration, sensors, etc.

At the core of a human-robot collaborative strategy is this deictic references of 'this' and 'that.' These are the references that are common to the representation the human and robot share. This thesis would like to add to the question of "where is the conceptual interface?", the more specific question of "how do you make common reference?" This particular question is addressed in this thesis through work with a virtual model. This is the problem of how

[8] A 19th century vision of body-brain interaction that separates each out into separate components, this view as a scientific explanation of human functioning shows how persistent the ideas of a philosophy that separates mind-body is, e.g. from Rene' Descartes.

can the human and the robot share reference for task objects which is difficult when the operator is removed from the task space.

Pook's work is a good example of engineering a solid interface between the robot and human and championing the cause of the the human robot interface in general [86]. The work involved some trial tests of the system with users, however these were primarily tests of the robustness of the visual recognition system as opposed to a study of the design of the interface. The work broke some new ground and the hope is that it will be picked up by others.

2.8.3 MissionLab Toolset

The MissionLab toolset is a graphical system intended to supply a human user with high-level control of a robot, or set of robots. This high-level control is at the level of the "mission." The mission is defined as task specification in combination with the ability to monitor task progress and make corrective actions. Thus the purpose of the interface is to supply germane data to the human mission controller and allow appropriate control actions to be made through the robot system by the human user. This system is considered pertinent to this thesis discussion in two regards. First, that it is an interface intended to control semi-autonomous robots possibly located over a distance. Second that the system designers have undergone user studies with potential expert users of the system.

As mentioned in the discussion on teleoperation the authors of this system have integrated teleoperated control with behavior-based robotics, both as an operator conducting the behaviors and as an additional behavior in the overall system. This is a unique perspective on the integration of human and robot control. The points of collaboration are more implicit and the human influence may be more subtle than the direct approach of other frameworks. This discussion however focuses on the "usability" evaluation of their graphical system MissionLab [71].

The evaluation of the MissionLab was composed of a comparing the compiled programming language instructions in the Configuration Description Language (CDL) with the newly designed MissionLab graphical interface built on top of the CDL library. In the study, users created and executed the same plan by two different methods. A comparison was made between users who used CDL to program in C and then compile their coded plans for a specific task and and second group of users who used the MissionLab interface to graphically specify a plan for the same task. The equivalency of functionality of the two systems is stated in the description of the experiment. What the authors define as the focus in their usability study is a comparison of the time users took to design and compile a plan for execution by the robot. It is not clear however that this is a fair comparison or that the results will tell us much.

In the study the authors make a number of assumptions about the use and study of their system. Some of these are: efficiency is the only goal;

there are objective measures that can be made; and statistical mathematics is the only way to explore system design. Unfortunately what is lost in this study are a number of qualitative and design factors. Such factors include: a sense of the use of the system, what sort of improvements might be made, and what features are important in terms of engagement of the user. In the study "target levels," measured in time, are set up to be achieved by the interface. The sole goal of the design of the interface is to obtain these values. Such a table is used in order so that "the designer can focus his/her efforts on improving performance in areas that are important, instead of *wasting time* on improving *insignificant* aspects [emphasis added]." Since the table only defines target completion values for "some indeterminate task" that can be "measured by concrete performance metrics," one can assume that everything else lies in the realm of the insignificant and is a waste of time.

This absolute focus on efficiency is worrisome. Although it is a common view of the technical perspective on usability. It lacks any sensitivity to design or first principles of user experience and it may be losing sight of the intended goals of simply designing better systems. Efficiency and cost are important for any system, but it seems to be easy for one to hyper-focus on "objective measures" while offering no insight for the enterprise of designing human-robot systems. This system is brought into the discussion here as it is one of the few mobile robotic systems to undergo any sort of user study.

2.8.4 Multi-Agent Supervisory Control

The Multiple Agent Supervisory Control (MASC) system is a 2D based interface offering human supervisory control of multiple robots. This system is explored in depth as it is one of the few systems that has tried to examine "human factors" as they relate to semi-autonomous control. This system and its study is also used as a point of comparison for the style of study in this thesis. It shares a similar position on the relationship between the human user and that of the robot to this thesis. The human acts as a supervisor and controls a robot that has some autonomous behaviors. It is one of the few such systems which has undertaken the task of giving attention to the human-robot interface. Given that background, and the similarities of supervisory control to the system in this thesis it is not difficult to find differences that take on significance in the comparison. Below is a brief analysis of the user studies that have been performed on the MASC system. This analysis is provided to offer a comparison in viewing the present system.

In the the user study of the MASC system the agents are heterogeneous and are composed of four robots, two sensor robots and two manipulation robots. The study was composed of recording user actions, video taping, and a sequence of questionnaires and formulas to determine subjective qualities of the interface. It involved 13 users with over 150 trials [3, 2]. The user was given the task of moving the robots around to explore the environment and to

position the robots in specified locations.

Unfortunately the authors of this study missed an opportunity to explore the relationship themselves between the robot and the human supervisor and instead employed an outside consultant to conduct the user test. Rather than generate insights about this specific system and the specific users, the study became a presentation of statistics and mathematics that support the rigor of the figures. The authors never explored, as often is the case with such quantitative studies, what assumptions were employed in the interface design, what their significance is and how they held up in the study. Most of the study presentation is composed of dense statistical results focusing on efficiency issues such as time of completion in various modes. In the last third of the paper, there is an informal discussion eliciting some of the observations on the use of the system. The observations are often limited to simple statements lacking ensuing discussion which might provide both specific and general revealing points about the interface thus missing a chance to describe more about the experience of using the system.

Many of these comments explore the interface at a superficial level and make a number of assumptions about buttons and clicking that are never addressed. In general the tests seem not to transcend the world of point and click 2D interfaces (GUIs). Many assumptions seemed to be made about the world of interfaces which contain the GUI notion as a foundation, however these assumptions are never explored or justified.

Though it was stated that a "consultant" was retained to help design the studies, no further information was given about the consultant. There are a number of references in the text to experimental psychology texts, so it is assumed that the consultant employed methods and framework of that tradition. The abundance of statistics and presentation format, as well as "randomization of tasks" would seem to confirm this. It is not clear that the authors and designers of the system had an opinion on how the study was conducted or how it affected the results nor of what actual questions about their system they wanted to answer. A key goal of the study appeared to be validation of the system and methods. However the validation goals can be brought into question. The methods of the study forced a change in the content of interaction. The full range of control of the system was not employed in the tests. In justifying the limit to the task space for the experiments the following is stated:

> During the design we determined that if we wished the subjects to execute difficult tasks, the time and monetary requirements would be beyond our means. This difficulty level was associated with the overall multiagents systems design. The multiagents system is fairly complicated and would require extensive training concerning the mechanisms and processes involved. Also, the overall multiagents system is not sophisticated enough to execute difficult tasks.

The above seems to be contradictory. The system is both "fairly complicated" and "not sophisticated enough". There are a number of possible explanations for this contradiction: this domain is inherently difficult, a framework has not been found to formulate multiagent work, or the interface itself has not been well formulated. Although the first is true, it is not presented as such in the work and the other possibilities for this contradiction are not addressed. In addition, the subjects were limited to certain modes of operation because of "the immense training required to operate the system in the other system modes." The use of the word immense in this context is worrisome as regards the usability of the system. This is especially worrisome because the tasks used in the experiments were very primitive and were not in effect a fair test of the system. Because of the statistical requirements of the experiment setup and the above mentioned "complexities" in the interface, task control was reduced to basic manually controlled navigation. Even though behaviors for obstacle avoidance were available they could not be used:

> The pared down version permitted the subjects to use all four agents and their sensing modalities. The locomotion command generation method was teleoperation and the autonomous locomotion methods were not employed.

The tasks were at the lowest level possible: direct control of the robot steering mechanisms. One would have guessed that turning on the autonomous control would make the system more independent and easier to use, not the opposite. It is hard to perform experiments, and this is partly why they are not done. Prototype systems are not always reliable and are hard to have working at a specific time, for example for users in a study. This was also true in the study in this thesis, the speech and arm sub-systems were unavailable for the study because of a number of complex issues involving system configurations, versions of software, and a mechanical breakdown. However, one should not allow a study method to undermine the actual system that is to be studied in effect, making an impotent system the subject of a validation experiment. Though the author claims that they have shown that users can communicate with the robots at the "task level", they have defined the task level to be so close to steering the robots, that it is not clear that they have in effect shown anything of the sort. Rather they show that a user can use the "complicated" interface to move a robot from one place to another via a 2D map.

Despite the statistical methods, for which the reduction in the task space was made, few cross-correlations were found. This may be one of the strongest criticisms of the work as it does not question the study itself but the results. One of the strongest correlations found was that between the number of commands issued and the complexity of the tasks, which also corresponded to a "workload measure" of the tasks. It is obvious that since the commands are issued at the locomotion level and thus increase with complexity of the task (a compound locomotion command) they are therefore more time consuming and

tedious to enter, *i.e.* two commands take longer to enter than one.

One study result that that provides some justification for the Reality Portal system in this thesis is the inadequacies they found when user's tried to use the video to navigate:

> "Another major difficulty is the agent's narrow view provided by the cameras. Many times the subjects and agents are unable to acquire a significant environmental view which would have assisted them in their task. We as humans have a 180deg field of view but when we explore what we can obtain from the agent's field of view; it is significantly less."

It is these types of limitations that the CVE plus textures seeks to alleviate.

Examining in detail the human factors study of the MASC system helped to solidify the design and approach of the study in this thesis. In the study of this thesis system more autonomous capabilities were demonstrated and employed by users than the MASC system, albeit in a single robot case. Users were also able to employ the system for a non-trivial task with a five minute demonstration of the system tools. In contrast to the "controlled-"style study above, the sessions were interactive and changes were made to the system to prototype new ideas and improve the system while the study was conducted. The focus of the study was the improvement of the interface, both through concrete design changes and through better understandings of its use.

2.9 Emergence of Service Robots

With the emergence of robots in the realm of the general citizen there needs to be a greater understanding of how these robots can be controlled and interfaced. This work has been carried out in part association with the Center for Autonomous system (CAS) at the Royal Institute of Technology in Stockholm (KTH). One of the principal projects in that center is that of developing a "Service Robot" that is constructed for use by the general citizen in a home use scenario. A living room environment has been created from standard IKEA room furnishing for experiments on testing the viability of perception and navigation algorithms in the "Real world" [7]. In addition, this program has launched a task to specifically explore the interface for such a service robot in cooperation with the Interaction and Presentation Laboratory (IPLab) at KTH. This move represents a progressive focus on the interface as vital area of research for robots that are to be used by humans [84].

The perspective taken in this thesis work draws on the broad number of communities mentioned in this chapter. It is believed that each has contributed important aspects to the design of such a system and it is through an integration of these works that the general topic of moving toward human-robot collaboration is explored.

Chapter 3

Approach: Human-Robot Collaboration

3.1 Introduction

The presentation in this chapter centers on the interface between a human and robot. The interface represents a connection between different spaces and competences. This chapter first offers a perspective on Human-Robot collaboration in general and then presents an overview of the solutions in this thesis.

In *Telerobotics, Automation, and Human Supervisory Control,* Thomas Sheridan presents an excellent introduction to interaction between humans and robots and the various models employed in the research. While Sheridan's presentation concentrates on the issues as viewed from the perspective of systems engineering, this chapter offers a complementary perspective which is on the interaction space and the channels of control. It is through a look on this space that one can then locate the solutions in this thesis. This chapter explores these issues, provides an overview of the implemented system and ends with a number of specific example questions that can then be asked once the system is in place.

The main claim of the thesis is that a 3D graphics system, in the form of a Collaborative Virtual Environment (CVE), provides a basis for supervisory control that is scalable relative to robot competence and both flexible and effective for remote environment visualization. Before looking at those particular solutions it is helpful to first explore human-robot collaboration over distance and the requirements for human-robot interaction and collaboration.

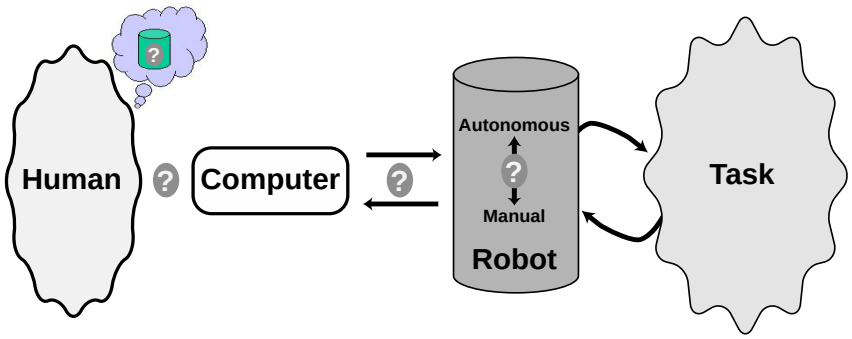

Figure 3.1: The basic components of a human-robot system. The question marks indicate areas of investigation. How does the human interact with the computer? What sort of medium of communication is there between robot and the human's workstation? How much autonomy does the robot have in the task space? In addition there are questions about the human's conception of the remote robot and system.

3.2 Toward Human Robot Collaboration

The applications for completely autonomous robots are many. However, the problem of autonomous mobile robot high-level task planning remains hard. A robot that performs complex tasks in unstructured environments requires high-level instruction from a human operator. For this a human-robot system for supervisory control can be constructed.

The illustration in figure 3.1 presents the basic components of any such human-robot system where the human and robot are separated by a barrier that prevents direct interaction (this might be distance, scale, or hazard). The question marks in the illustration point to where some of the basic questions about such a system lie. From right to left these are: 1). How much autonomy does the robot have to interact with the task environment? 2). How do the computer and the robot communicate and share information related to tasks? 3). How does the human interact and specify tasks through the computer and receive feedback from the robot? In addition there are other questions about the human's model of the robot and how their expectations and conceptions are brought to bear in the use of the interface.

In trying to implement solutions that answer these questions there are some basic requirements that need to be met. For the system to support human-robot supervisory control these requirements involve design decisions centering around the following needs:

- The human operator needs information on the physical surroundings of the robot.

- The operator needs to be given a reasonable rendition of the robot's current awareness of those surroundings.
- The operator needs to be given a useful and understandable mechanism to make joint remote environment references in order to be able to specify objects, entities, and tasks for the mobile robot.
- The operator should have some method to evaluate (and possibly negotiate solutions for) robot task successes and faults.

These needs form a system that enables a user to visualize and interact with remote environment, and interact and specify tasks with the remote robot. Looking back at figure 3.1 the next sections then explore the following questions:

- What are the modes of human-robot task interaction?
- What are the human-robot channels?
- What interface might support human-robot collaboration?

3.3 Modes of Human-Robot Task Interaction

The purpose in having a human robot interface is to enable a human to use the robot to perform a task. If *supervisory control* is defined broadly, as "that which is between human manual and automatic control" we are left a great deal of space in between manual and automatic for different styles of performing a task. In figure 3.2 there is a taxonomy of models that depict the different levels of interaction and autonomy between a human operator and a robot. On the left is the situation where the human has full control through the robot interface, on the right is the situation where the robot is fully autonomous and the human is in a monitoring role. This is a coarse model and there are many degrees in-between. It is these degrees of autonomy that are referred to as the scale of robot competence. A good system for human-robot interaction would tolerate shifts in robot autonomy on this scale without major changes to the interface.

One aspect that the taxonomy does not address is how this work is divided given a hardware framework (*e.g.* given the connections in figure 3.2). There is another taxonomy that, divides up the activity into a number of styles of work for performing a task. These styles of division are referred to here as *modes* of interaction. They describe, by invoking comparisons to everyday work practice, the ways in which we might share task work with a robot.

What are these modes? Sheridan has invoked the concepts of *sharing* and *trading* as distinct modes of interaction. These are further broken down into the sub-categories of extend, relieve, back-up and replace:

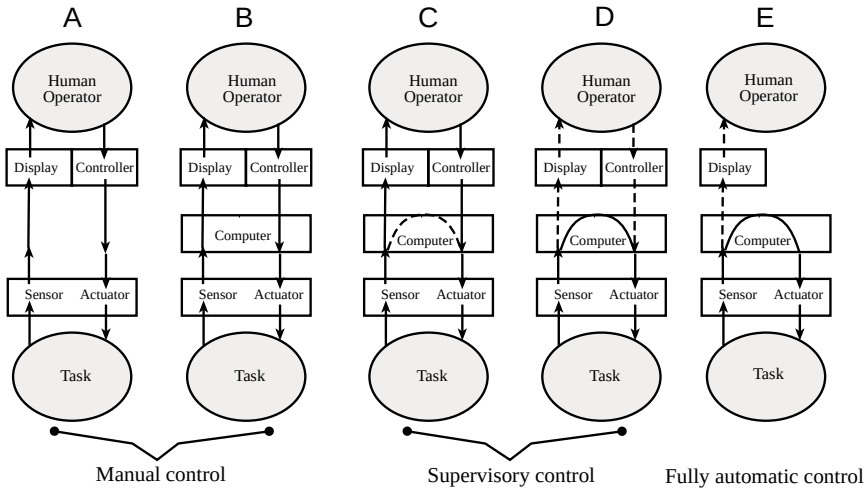

Figure 3.2: This figure shows the different degrees of robot autonomy. *A* and *B* represent human manual control with no or little computer intervention. Figures *C* and *D* represent supervisory control where there exists a closed control loop under computer control. *E* represents a fully autonomous situation where the human plays a monitoring role. Dotted lines indicate minor loops and solid lines major loops. The figure is adapted from Sheridan[92].

> The computer can *extend* the human's capabilities beyond what she can achieve alone, it can partially *relieve* the human, making her job easier; it can *back up* the operator in cases he falters; and it can *replace* her completely [92].

These modes locate the human and the robot in a number of situations where the division of work is decided and deliberate. Another category exists where the division of work in tasks may not neatly divide *a priori*. Such situations are when the task is such that decisions need to be taken while the work is in progress. The next step of the process, or the next action, may depend on the results obtained by the previous. For such tasks it would be difficult to make a division of the work beforehand. Another such situation is in dynamic environments where the constraints and perhaps locations of the task are changing frequently. There are also sure to be tasks where the ability to perform an action is impeded. This might be caused by an impediment such as lack of competence, lack of information, or lack of dexterity, etc. In these situations, although it might be possible to divide the work of the task beforehand, the preconditions upon which that division was made have changed. This category of situations indicate a need for a *collaborative* mode where the division of labor between the robot and the human is not set, but can be fluid. It could

be claimed that any complex task in unstructured environments would involve this sort of fluidity between modes. It is this mode of work that this thesis refers to with the term *human-robot collaboration.*

Working division of labor This fluidity of the division of labor is referred to as a *working division of labor,* that is the division of labor is a working one (see section 2.7). With such a working division, elements of the task might then run the gamut of modes: extending, relieving, backing-up, and replacing. In fact, these swapping of modes may be entirely tacit, *i.e.* happen without conscious attention. The shift between these modes may happen without clear boundaries and thus be inaccessible to techniques such as task analysis or other such set divisions. There needs to be 'working space' for human and robot to attend to the moment-by-moment contingencies that emerge in the doing of work. Work in CSCW makes the case that this is so in the use of computer systems, and as stated in the earlier chapter, this thesis claims this is also a situation with the use of robots.

As regards the interface, to allocate roles and resources (*e.g.* between human and robot) on the basis of some formalized notion of a plan-in-advance, ignores the practical realities of situated action and therefore risks "designing out" support for the realities of effective collaborative work. For the practical design of the interface this implies a construction that enables the sharing of tasks where this facility is part of the base structure of the interface. An interface can provide for this not by just one supported method for direct manipulation but by providing several options available *at any time* enabling the human and robot to work together and collaborate.

Semi-autonomy as a mode A semi-autonomous robot is dependent on the collaborative relationship with the human operator and this relationship is vital to the working of the system. One need is to provide the robot with primitive basic behaviors. These basic behaviors can then be combined at the interface level to perform more complex tasks. The robot has some given basic competences and then the interface and the human user are able to compound these into more complex commands. To look at a specific example a robot has behaviors consisting of: 'avoid-obstacles,' 'goto-location', 'move-forward,' 'turn-around,' 'look there,' 'grasp-object', 'release-object'. A more complex command can be given at the interface level such as: 'go-there-and-look.' It is through these behaviors in combination within a particular interface that a human operator interacts and collaborates with a robot and that level of semi-autonomy can rise or fall. Here semi-autonomy is a mode of working with the robot.

3.4 Channels for Human-Robot Interaction

What are the requirements for Human-robot interaction to take place at all? One of the first requirements is that there is a purpose to the interaction. This thesis has considered the situation where a human user has a desire to perform a task in the physical location of the robot but where the robot and human are not necessary co-located. Given that as an assumption, what are the available channels? For the human to control the robot, at any level, there needs to be a form of communication. The human user needs to be able to supply information to the robot which can then be interpreted as control commands. If the robot is in a dynamic environment, or in one which the task depends on the configuration of the physical space, it would then necessary to offer the human some way of understanding that space.

3.4.1 Visualizing the environment

In most remote robot systems, the channel used to visualize the remote environment has been, in fact, visual. Some form of visualization is provided to the human operator in order for the human to be able to asses the needs for the task.

For a primitive system, a form of communication from the human to the robot, and a form of environmental feedback from the robot to the human would be minimal requirements. If we consider the task of picking and placing objects in a remote environment, where the selection of objects depends on visual detail, we could then consider some form of remote video to be used for the visualization. Such a remote system could be commanded to roam around the remote space and send back continuous images (*e.g.* 25fps) offering the human a way to visualize the space. These systems where a camera is positioned on a mobile robot and where the video is sent back to the robot are not uncommon. Such systems do however have limitations. The limitations are due to difficulties in storing and reviewing previous images as well as the problems of being coupled to the dexterity of the robot. It is hard to see what is outside the video at any one time and hard to look quickly to the sides. Linking the human closely to the mechanics of the robot may solve some of these limitations but introduces others such as lag, which may never be overcome due to communication delays.

3.4.2 Interacting with the robot

In *Introduction to Robotics, Mechanics and Control*, Craig makes the following point about the difference between the way positions in a robot's mechanical space are calculated and the way a human can be asked to specify positional parameters.

> In order to make the description of manipulator motion for a human user of a robot system, the user should not be required to write down complicated functions of space and time to specify the task [33].

References to space and time are more comfortably specified in space and time. The "learn by showing" school of robot programming demonstrates this exact difference over the older method of "pendant" robot control. It is also this kind of distinction that is addressed in much of the human robot interface literature that has used virtual environments or other GUI-style interfaces described in the previous chapter. A human user needs to be able to refer to parts of the remote environment. In particular, task specification requires the interface to support deictic references. These are interfaces that enable references through concepts such as "that," "there," or "this." There then needs to be a medium for this communication to enable the references to be part of task specification and interaction.

Semi-autonomy as a channel In Craig's statement there is a notion of autonomy on the part of the robot. Tell it where to go, or what to manipulate, and let the robot perform the calculations and functions necessary to perform the action. With the concept of semi-autonomy, some traditional critical factors and research problems in teleoperation, such as time lag and feedback loose their relevance. Lag, the time between when a command is given and when it is acted on by the robot has been well studied. That study has had as its goal the reduction of lag, as well as to discover through psychophysical experiments, how much lag may be tolerable to a human operator before lag becomes an issue of distraction to the task at hand and in some situations, lag is inevitable. Such systems may seek to unify the robot and the human where the robot becomes the direct remote agent of the human controller. In contrast, if the system is set up with semi-autonomy, *e.g.* within a supervisor-assistant system, much of the weight of this issue is diffused. By offering higher-level commands to the more competent robot the need for feedback becomes less frequent. Here semi-autonomy is a channel, the required bandwidth of which can increase or decrease depending on the level of autonomy available.

3.5 A System to Support Human-Robot Collaboration

Through the CVE a connection is made between a human operator and a robot partner. Using the CVE the human can visualize that remote space and specify deictic commands within a common reference frame. In figure 3.3, the CVE as that common reference frame is made explicit. The CVE is the communication medium, a shared distributed database, between the robot and the human supervisor.

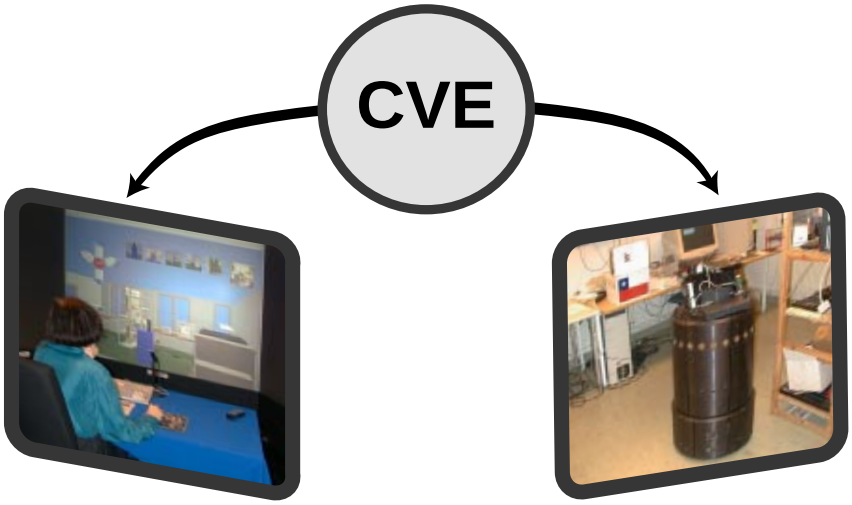

Figure 3.3: The human on the left collaborates with the robot on a task by using a CVE to visualize the remote environment and interact with the robot in the task environment via the virtual model.

The virtual environment framework combined with semi-autonomy attempt to address many of these elements. In figure 3.3 the CVE is the connection that provides the framework to answer questions 2 and 3 from the discussion of figure 3.1 in section 3.2. Through this framework the human can interact with the semi-autonomous robot that shifts its competence on the scale of autonomous vs. manual control.

The main components of the approach in this are the following:

- Supervisory Collaborative Framework
- Environment Visualization
- Interactive Control

These are presented in the following sections.

3.5.1 Supervisory collaborative framework

The DIVE virtual environment system provides the CVE platform upon which the application system is built. The graphical model runs on this platform and represents the structure of the remote space. The CVE provides mechanisms for communication and the sharing of 3D object and function database. These mechanisms provide the support for communication between the different computers distributed over the different sites.

Within this spatial environment, concepts such as *robot context, relative position, dynamic behavior* are available by the nature of the dynamic interactive 3D virtual environment. By robot context, and relative position what is meant is not only the configuration of the robot in some coordinate system, but the robot location relative to other objects in the environment as well as the affordances of that environment and how the robot is situated to take advantage of them. An example of such an affordance relative to the supervisor, is that a book on a table affords being grasped by the robot, or rather affords a deictic command to the robot to grasp. By dynamic behavior, what is meant is the understanding of the robot motions that happen as the robot is observed by the supervisor. Over time the supervisor will have their own expectations of robot behavior based on past motions and actions in general. It is this working spatial sense of the robot environment that the 3D system offers over more traditional 2D interfaces with windows, panels, buttons, and strictly planer projections.

The physical robot has basic behavioral competences that run on its local on-board computers. These center around maintaining a positional map of the environment, navigational target maintenance and employing distance sensors to avoid sensed obstacles. As an alternative to treating human intervention as a special case, human-robot interaction is part of the guiding metaphor of a supervisor-assistant relationship. This metaphor also allows for growing the competence of the robot sub-systems without breaking the interface concept. As the competence of the robot improves the expected effect is reduced effort on the operator and a more expressive and competent system as a whole. The robot competence is centered on navigational and low-level perceptual skills while the operator is responsible for high-level control.

3.5.2 Environment visualization

The ability to explore the remote space is provided through the construction of an interactive augmented virtuality. The augmented virtual world contains real world images as object textures. This enables a user to explore a virtual representation of a real space. The textures are taken from objects that exist in the real world and which have dual (mirror) objects in the virtual world. This has the advantage of making a virtual world a representation of the real world, while maintaining the flexibility of the virtual world. Objects having the images of their real counterparts but can be manipulated in a virtual setting. The virtual world can also be used as a control interface for manipulating objects in the real world via the robot. An advantage of the virtual world is that it is not dependent on physical location and can be manipulated in a way not subject to the temporal, spatial, and physical constraints of the real world. It also has the advantage that irrelevant parts of the real world can be left out. Thus the "interesting" parts of the world can be extracted and made salient. In this way a custom view of the world can be created, forming an instantiation

of selective vision. The texture based virtual objects are referred to as *Reality Portals*. This system is meant to overcome some of the temporal and spatial limitations of having only a video image. In the technical work, a number of various methods of inserting video into a CVE are explored along with their limitations and reasons for having the Reality Portal method.

3.5.3 Interactive control

A main theme here is the provision of a supervisory system that enables human-robot collaboration. Within a collaborative relationship between human partners, the notion of a *working division of labor* has been identified, a relationship where the division of labor is not necessarily strict, but can shift, as appropriate. Mechanisms for control also need to support this shift. The methods and mechanisms for interaction include the direct control, gesture control and the speech interface. Inherently a 3D interface offers spatial interaction. In the case of a deictic system for robot control, this spatiality is at the core of the human-robot exchange. The CVE as the medium of communication between the human supervisor and the robot assistant can be seen as positioned in between the world of the supervisor and that of the robot, offering a bridge for a common exchange about the task-world in question. The CVE, as medium, offers a method of visualizing the robot task-space as well as an arena for interaction between the human supervisor and robot assistant partners in the task.

3.6 Technical work and evaluation

The next chapter describes the technical construction of such a system for supervisory control. Given the construction of such a system, specific questions can then be asked through its use, for example:

- What is the appropriate fidelity for modeling the remote environment?
- What are the effective metaphors for displaying remote video?
- What are the factors that affect trust and presence?
- What are the available interaction controls?
- What are the appropriate levels of command for a task?
- How can close interactive collaboration be supported?

This chapter has presented the general approach and motivation for the system in the thesis. The following chapter presents the technical aspects of the constructed system. Chapter 6 and 7 present the details of a study that posed some of the above design questions, as well as discovered new ones, through a Study of Use.

Chapter 4

Method: The Human-Robot System

4.1 Introduction

This chapter describes the technical work for constructing the supervisory teleoperated robot system in this thesis. What has been created is a working system that has been used and demonstrated for a number of years through demonstrations and a Study of use with users. The system as a whole is complex and composed of a number of sub-systems. Each of these systems presents its own challenges and solutions. For the purposes of this thesis presentation, the sub-systems can be classified as belonging to the categories of **Supervisory Framework, Environment Visualization** and **Interactive Control.** These categories are described below:

Supervisory Collaborative Framework: Providing the foundation.

CVE platform and model: The DIVE virtual environment system provides the Collaborative Virtual Environment (CVE) platform upon which the application system is built. The graphical model runs on this platform and represents the structure of the remote space.

Virtual Robot Agent: An interactive model of the robot is added to the virtual model of the space and contains the functionality required to communicate with the robot and the CVE environment.

Robot semi-autonomy: The physical robot has some basic behavioral competences that run on its local on-board computers. These center around maintaining a positional map of the environment, navigational target maintenance and employing distance sensors to avoid sensed obstacles.

Environment Visualization: Methods for visualization.

- **Video in the CVE:** The elements of hardware and code necessary to display video from the robot in the CVE.
- **Reality Portals:** The system that automatically selects and crops pieces of video to be inserted into the CVE.

Interactive Control: Methods for interaction.

- **Command interaction:** The set of graphical models and functions that provide the functionality of specifying robot tasks from mouse based input in the CVE.
- **Spatial Model:** The code and virtual models needed to implement the functionalities of spatially programmed functions.
- **Speech Interface:** The system used to provide speech input and output to and from the robot.

The robot assistant and virtual robot agent navigate and interact in two different worlds, the real world and the virtual world. The robot physically exists in the real world and the virtual world contains a representation of the physical environment and the virtual robot agent. The virtual environment is built up from an architectural drawing of the basic physical world structure and artifacts. Through movement and exploration, the robot has the ability to display video of the remote environment to the user through the virtual world and as the robot explores the real world it can augment this model with textures of objects taken from the robot video. Through the virtual world the user is also able to interact with the robot, specifying navigational targets, and places to direct camera focus.

To understand these sub-systems better, one way of viewing the system is as an information flow (figure 4.1). Information passed around can be seen as flowing between the real and virtual worlds via the camera, the robot and the user. The video from the camera flows from the real world to the virtual world. The video from the real world originates from a robot-centered perspective (*e.g.* an on-board robot camera) and is displayed in the CVE. The user issues commands to the robot via the virtual environment interface.

In this way the virtual environment can be understood as the communication medium between the robot and the user. It is through the interaction of the human with the CVE and through the interaction with the robot with the CVE that interaction between the operator and the robot takes place. The CVE is the medium for bi-directional communication, environment visualization and command specification.

These three categories and their sub-systems are described in the following sections.

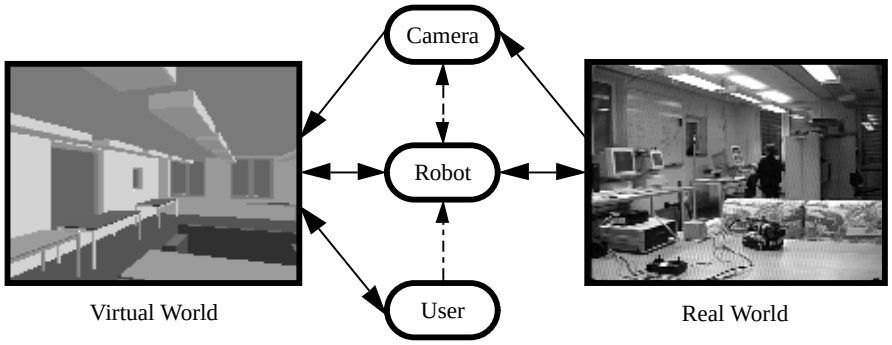

Figure 4.1: The figure shows the flow of information between the different system components. The solid lines indicate direct flow of information while the dotted lines indicate an indirect flow. *(image: Frécon)*

4.2 Supervisory Collaborative Framework

The supervisory framework provides the infrastructure for the implementation of the system. Within this framework there are three main distributed computational systems. These are *the CVE platform with the virtual model, the virtual robot agent*, and *the physical robot system and its semi-autonomous behaviors.*

The CVE infrastructure is based on the SICS Distributed Interactive Virtual Reality (DIVE) system [30, 50]. This system provides a platform for which CVE applications can be developed. This platform consists of a number of programming models (C, OZ, Tcl, Java), virtual object modeling languages (VRML, VR, AC3D), and the networking infrastructure to connect up distributed applications running on an ethernet (*e.g.* TCP/IP, ATM, URL, MIME).

The robot platform as provided by the RWI robot manufacturer is supplied with a set of programs and libraries known as Rhino. This platform provides the infrastructure for gathering sonar and infrared readings from the robot sensors, for controlling the robot motors and for reading encoder values. On top of that are a number of applications that run to maintain different robot functions. Through the TCX software interface to the hardware bus system, applications can be written that provides access to the robot hardware as well as enables communication across systems to make remote procedure calls.

This supervisory system in the thesis consists of a number of applications in both the DIVE platform and the Rhino robot system that are run on a host of computers on an ethernet network.

4.2.1 CVE platform and model

The CVE platform and model consists of the application programming platform of DIVE with the virtual model of the remote space. Together these provide:

- Platform for distributed virtual environments.
- CVE robot application platform.
- Visual Display of virtual environment.
- Method for subjective viewpoints, *e.g. avatars.*

The DIVE system at its core is a system for maintaining a distributed multimedia database on a network. Most applications built on DIVE use this database to visualize 3D graphical collaborative virtual environments. These environments can be composed of 3D models, sound, programmed behaviors, video, etc., but do not have to be composed of all. The models are generally specified in the VR file format, which can also contain TCL programs attached to 3D objects.

The applications that implement the event capturing and interaction with the world, the communication with the robot, and the spatial model, are programmed in a mixture of C and TCL. The communication layer with the robot is a DIVE application written in C that starts a TCP/IP socket with the robot and initiates a C/TCL interface process in the DIVE database that enables a number of the C functions for robot communication and interaction to be accessed through the TCL scripting language.

The CVE model of the environment is visualized through a pre-existing application called VISHNU. This application is part of the standard DIVE distribution and is run in order to visualize the 3D distributed database at a particular host on the distributed CVE network. It is this application that enables navigation through the virtual environment, and monitors interaction by the user to trigger events for other applications to monitor and act upon. The model that this program visualizes is specified before in a model file. Here the model is a CAM style model built in the VR file format based on measurements of the room. These measurements were made with a tape measure and are expected to be within a few centimeters accuracy.

In the CVE a number of users can share a virtual environment model. Often this involves a number of users to be stationed at distributed workstations yet be able to interact through the virtual model. Each user has their own subjective view on the model that is determined by a number of factors such as cartesian coordinate in the environment, orientation, angle of view, and a host of other custom features. The coordinates and angle are determined by a user's six degree of freedom (6DOF) position in the virtual space, other factors such as viewing and angle and rendering options are determined in the setup of VISHNU. The user 6DOF configuration in the virtual environment is marked

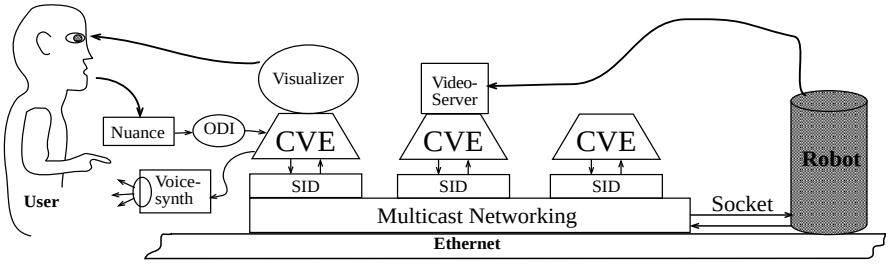

Figure 4.2: Overview of system architecture with components focusing on the CVE core of the system.

by an *avatar*. An avatar is the user's representation in the space. An avatar can be seen by other users in the space, and attached to this avatar is a *visor* where objects can be attached to the user and remain co-located even as the user navigates the space.

Thus the virtual world can be seen by the user to represent the "knowledge state" of the robot, and to visualize the robot's configuration in the environment. This is not intended to be perfect or complete, but to be enough to perform the task (this point is explored further in the Study and Findings chapters). Because this representation of the robot and environment is visual it gives the operator instant access to the robot's current situation. The next section describes the robot agent in the virtual world.

4.2.2 Virtual Robot Agent

The robot has a representation, or avatar, in the virtual environment. This representation is referred to as the *virtual robot agent* to distinguish it from the physical robot assistant. The representation can be seen in figure 4.3 and is intended to provide the following:

- A depiction of robot configuration as a virtual model in a 3D CVE;
- An embodiment of communication methods with the robot.

The provision of the virtual robot agent enables a clear way for a supervising user to visualize the robot configuration relative to the space of the virtual model. Because this virtual model is a representation of the real physical space and because the virtual robot agent's position is updated at about 10hz from the physical robot's localization system, this virtual representation enables a user to understand the physical robot's configuration in the physical space.

The virtual robot agent's position is updated as the physical robot moves and thus has dynamic behaviors in the virtual world. There is measurable lag in these movements, but for the most part these do not play a role as the user

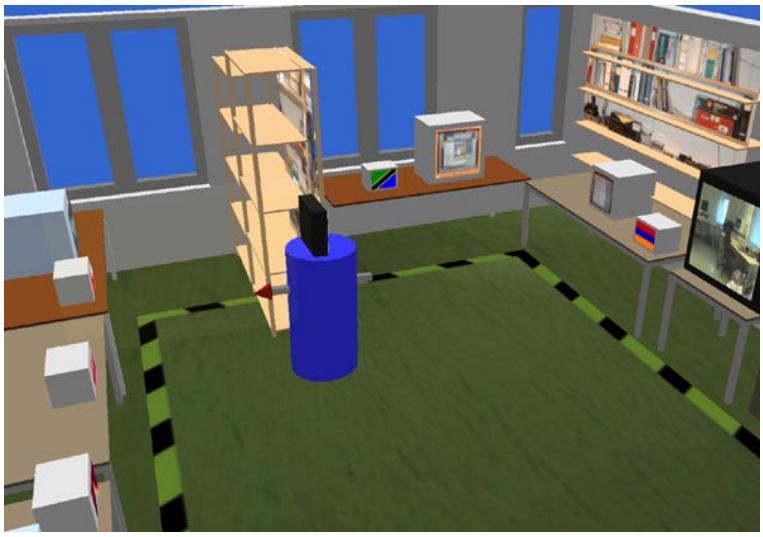

Figure 4.3: The figure displays a view of the virtual robot agent situated within the 3D model of the remote robot environment (see also color figure **C**).

is primarily working with the robot through the medium of the CVE. When video of the physical robot is displayed in real time to a user this can cause confusion (this is explored further in the next chapters).

Anthropomorphism

A design decision has been made with the virtual robot agent that it not be made too "lifelike." By lifelike, what is meant are those features, or behaviors, that make the agent more anthropomorphic, (*e.g.* eyes, mouth, expressions). Although the technical capability was readily available, such anthropomorphic features were not necessary in this interface. In the DIVERSE project also based on DIVE, such a virtual agent (virtual Ray Charles) was made available to represent the speech systems competence [60]. From an interface design standpoint this capability was not appropriate for interacting with the robot. This is for a number of design reasons: interface transparency, tacit collaboration and anthropomorphism.

The interface is actually composed of at least two levels, that of interacting with the virtual environment and that of interacting with the robot. In this system the physical robot can be seen as the agent to be interacting with. Thus a virtual agent in between would not only be confusing, but put an extra level in-between the user and the physical robot, reducing the transparency of the interface.

It is a design goal of this system to realize a working division of labor.

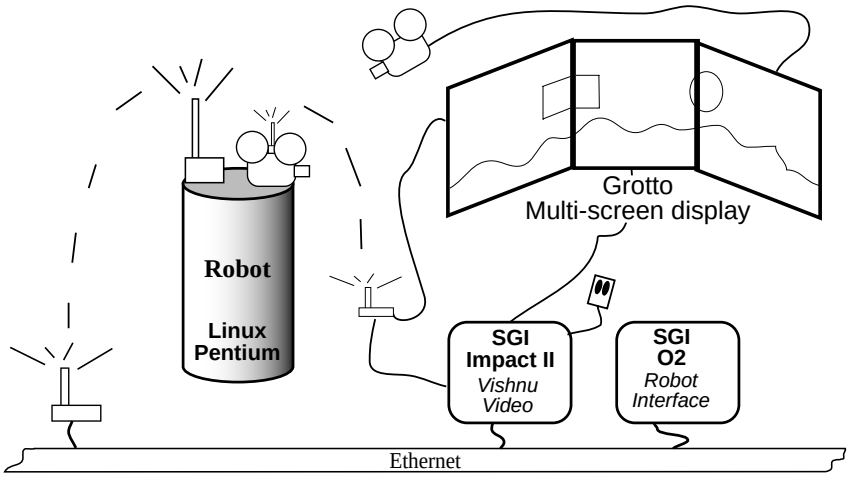

Figure 4.4: The image depicts the salient components of the hardware system to run the system as used in the study chapter. The robot runs untethered with radio ethernet and radio video connections. A number of Silicon Graphics machines are situated on an ethernet, and video and graphics for the CVE drive the multi-screen Grotto displays.

One way to approach this is by making collaboration a tacit feature of the interface. That is to say no "lock-step" methods are required to implement collaborative communication to the physical robot. Such virtual agents tend to make collaboration explicit by agreement. In previous versions of the virtual robot agent system, the robot spoke to acknowledge commands. This was deemed inappropriate, even though in the study it is revealed that the proper approach here is a middle way.

Third, anthropomorphism of an interface needs to be treated with care. As mentioned in the discussion on attributes and endowments in Chapter 2, over-anthropomorphism can create unrealistic expectations on the part of the user. Unfortunately these sorts of interfaces techniques are adapted too easily without considering the downsides of their use. If there is to be some form of anthropomorphism of the physical robot or virtual agent it will be because it is determined to be necessary for a particular task.

Framework hardware and software setup The framework is complex, employing multiple computers and heterogeneous hardware and software platform architectures. This complexity is outlined to the level of detail of where changes in the architecture made a difference to the system and where changes or modifications were made during system construction. Where such components where part of the standard system configuration (*e.g.* the IP networking

layer of the SGI computer) they are left out as basic portion of the assumed infrastructure. There have been many versions of these system architectures as the system and the systems it depends on have evolved. What is given here in figure 4.4 is the latest architecture that represents the system as it stood for the Study chapter of this thesis.

4.2.3 Robot semi-autonomy

The intention is to present the robot and its competence in as honest a way as possible to users. By presenting the system as explicitly "semi-autonomous" there is a contract based on the robot limitations made with the user. This manner of presenting and making explicit the limitations of the robot is meant to keep expectations on the part of the user relative to the robot's genuine competence. Part of this contract is the search for appropriate terminology and another part is the "designed with care" presentation of robot ability. This also involves more subtle features such as the possibility for interaction and the user to "take control" when the robot is seeming not to fulfill a specified command (this is discussed further in the interaction section). What this translates to for the on-board robot competences is the provision of command preemption in the communication interface and behavior execution environment on board the robot. That is robot actions can be stopped, redefined, launched by the user at almost any time.

The robot's basic competence includes point-to-point navigation, position estimation and the ability to avoid basic obstacles in the indoor structured environment in which it is situated. The navigation and obstacle avoidance are discussed in more detail below. For position estimation the robot currently uses an initial position together with its on-board positional encoders and navigates by "dead-reckoning." The encoder on the RWI B21 robot base is quite good and although there is drift, it was insufficient to cause concern during the sessions the robot has been run (up to one hour). A ready extension to this dead-reckoning is the incorporation of methods based on the author's previous self-localization research [104].

Navigation and Obstacle avoidance

The basic semi-autonomous competence the robot has are routines to perform the "point-to-point" navigation. In addition to this the robot has the ability to avoid encountered obstacle and make local corrections to its path. Specifically this competence takes the form of goal-oriented navigation with basic collision avoidance of sensor-registered fixed and moving obstacles that it detects. It is this basic competence which the operator actuates when a location is specified as a goal point for point-to-point navigation. These navigational target points are sent to the robot through the communication socket by the virtual robot agent. It is these competences which are made explicit and brought out

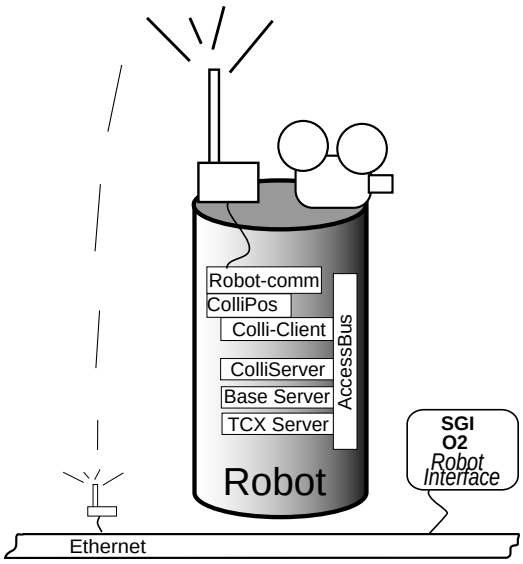

Figure 4.5: The image depicts the components of the robot software system to run the system as it existed in the study chapter. The robot runs a number of levels of programs on its Linux Pentium computer which handle collision detection, navigation, and communication with the CVE robot interface application.

both in the description of the system (*e.g.* as "semi-autonomous") and in the explanation of the systems limitations to users.

The goal-oriented navigation and obstacle avoidance behavior is based on the BeeSoft ColliServer subsystem developed at the University of Bonn and Carnegie Melon University [47]. The ColliServer is a dynamic modeling of the environment based on a small temporal window of sensor readings and of the robot's dynamic configuration. Inherent to decisions about navigation and avoidance are models of the robots velocity and accelerations in determining the ability to react to perceived obstacles and avoid collision while maintaining goal-directed behavior. The navigation system, which steers toward target points, and the ColliServer, which steers away from obstacle, run in parallel. The ColliServer sets temporary high priority targets that steer the robot away from obstacles. Such a system has proved to be a rather robust base for use during a longer period of time in a cluttered and dynamic environment (along with a higher-level planner and mapping system) in the Deutsches Museum-Bonn in 1998 [28].

In its implementation in this system the core of ColliServer has been maintained while updating for the specific differences of the physical robot "Rose-

bud" used in the experiments of the following chapter. The systems interaction with the higher level has been altered and adapted to be used with the DIVE-based virtual robot agent. This alteration concerns primarily two sub-systems that link the different coordinate-spaces of the robot and the virtual environment and the communication across different platforms. Specific functions have been developed that rely on the ColliServer core functionality that allow the integrated transformation of a CVE-specified point-to-point navigation command into the activity of ColliServer goal-based navigation. In addition a number of previously undiscovered "bugs" in the ColliServer were located and repaired to make it work with the current system. These bugs did not appear in the German Museum application because it is believed the sections of code employed here were not functional when the system is using the higher level Rhino Planner and Mapper. This is to say that although the core system and code are ColliServer as distributed (unsupported) from Real World Interface (RWI) robotics and the Rhino project at University Bonn, the system has been significantly customized and updated for this work in its current configuration. The levels of software on the robot are outlined in figure 4.5.

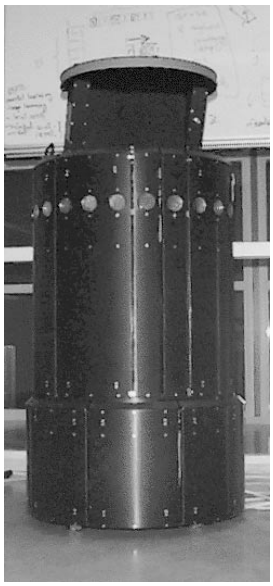

Figure 4.6: Frontal view of real robot. It is a Real World Interfaces B21 robot with mobile base and an on-board Pentium Unix-based processor. It has a number of sensing systems including a ring of sonar range and infrared proximity sensors as well as integrated contact bumpers all of which are employed in obstacle avoidance navigation.

Physical robot The robot used in this system is a Real World Interface B21 robot with on-board processing and sensing (figure 4.6). The robot is powered by four sealed motorcycle 12v lead-acid batteries. The robot contains a custom built power-management system that supplies on-board and untethered power for the robot to roam for approximately 60 minutes on a single charge (this of course depends on motor usage). The robot also has twenty-four sonar and IR sensors mounted on its enclosure delivering a compliment of range measurements of both near (IR) and relatively far (Sonar, up to approximately 25 feet) targets. The robot has mounting cages for two PC units in the upper enclosure of the robot. The robot used for experiments in this thesis is, named Rosebud, contains one such PC that runs a version of Linux 2 OS software. On top of the OS layer is a software infrastructure that enables basic functioning of the robot and control of its sensing and actuating systems. In addition there is a neatly designed gripper stalled inside the bay that can be deployed for pick-up tasks. The robot connects to the network via radio-based ethernet and transmits its video to a fixed station via pair of consumer video radio links. Both of these transmit in the 2.4GHz range.

Low-production, experimental and complex systems

This robot software architecture is built and quasi-supported by a cooperation between RWI and the University of Bonn. The software system is distributed as BeeSoft and is based on the U. Bonn system RAI. This system can be fragile with respect to changes, but if setup carefully can provide a solid foundation for which to implement user-created robot control software. There are few such robot systems in existence and the fundamental hardware platform is liable to change, making the next releases of the software platform liable to cease working with changes in hardware components. When the hardware and software is managed carefully and maintained relatively statically (*e.g.* rejecting unnecessary upgrades) the robot system can provide a solid development platform. However this often means not having access to the latest peripherals (*e.g.* wireless ethernet cards) as they require an upgrade to certain versions of the OS that may not be yet supported by the robot software systems. This also means that changes in any of the sub-systems can cause other parts to cease functioning.

The system is also sensitive to hardware peculiarities that are common to low-production custom hardware. The most common problem in this realm is related to the power sub-system. Often a symptom of a malfunctioning disk-drive, which could be interpreted as caused by a need for reformatting (which in fact may be required), or a faulty ethernet card, may have as its original cause a fluctuating power supply. At this stage in the B21 robot hardware development, such problems mostly occur as the result of aging batteries in need of replacement. Fundamental behavioral differences can occur in high-level software when on battery power versus when plugged into the wall. As

stated these problems have mostly been rooted out, but as Rosebud was one of the first handful of B21 robots made, these problems caused many hours of circuit board replacement and upgrades in the beginning.

Working with physical robots is a challenge in many respects and occasionally these problems can shift the focus of intended research work (*e.g.* the planned use of the robot arm). Much of this is due to the challenge of working in a dynamic "real world" environment as mentioned in the discussion on software agents versus physical robots. However this is also the result of working on a custom platform that has not undergone the sorts of rigorous testing that a standard consumer platform might have undergone. Generally these units are sold as "research platforms," although, as stated earlier, this trend is changing as applications areas are opening up for the general citizen and more robots are becoming available as "off the shelf" consumer products.

4.3 Environment Visualization

The previous chapter motivated and outlined the need for environment visualization when attempting to collaborate with a robot on a task. The low model fidelity of the virtual environment and the desire for a robot to explore dynamic remote environments suggests that a graphical virtual environment model can benefit by the addition of video images from the real scene. The CVE model cannot supply all detail of the remote environment and definitely cannot supply information about those features of the environment that are dynamic. However the CVE model does contain information about the basic structure of the remote environment. The model then is seen as providing a base for the addition of textures from the remote environment.

How the video is treated, its source (*e.g.* camera placement) and its placement in the CVE affects the manner in which the user can visualize the environment. These differences can be seen as falling into two categories:

Metaphor What metaphors are used to convey the visual information.

Method What methods are used to construct and place the visual information.

The investigation in remote environment visualization has explored several techniques. *Reality Portals*, last in this section, is the ultimate of these visualization techniques. Prior to discovering the metaphor and methods for Reality Portals we explored other techniques that have been called the *Monitor Metaphor* and *Projection Screens*. Because the initial experiments led us to discover the final method, these earlier techniques are first explored in the following sections. The discussion then offers the limitations of these early methods and offers some motivations for the use of Reality Portals. Finally, the Reality Portals technique is presented as the most mature and elegant

solution to the limitations encountered with the other techniques. The next chapter offers a exploration of how users have perceived the system and the display techniques.

4.3.1 Video in the CVE

The physical robot has a fixed on-board camera pointed in its forward direction of travel. This camera sends video by a radio link to an SGI machine that is able to take it and incorporate it into the virtual world. Over the course of this robot work, the methods for displaying video into a DIVE CVE have been integrated into the VISHNU application described early[1]. Thus video has become a basic part of the VISHNU application environment. Unfortunately, because of the bandwidth restrictions of distribution, the frame-rate and maximum resolution of the video images is limited. Generally this video can be displayed at 128x128 and 15 fps inside the distributed virtual environment. Though this limitation affects the user perception of the video it is not a motivation for the solutions here. This video infrastructure is taken as a starting point for the exploration with the expectation that we have access to greater resolution and frame-rate.

The question has been how to place this video from the robot into the virtual environment so that it "makes sense" to a user looking at it. The first attempt at this has been the Monitor metaphor, placing the video stream on a graphical object in about the same place as the camera sits on the robot.

Monitor Metaphor

The *monitor metaphor* describes the scenario where live video from the robot's working environment is presented through a virtual monitor on top of the robot model in the virtual environment (figure 4.7). The live video presented on the monitor originates from a camera located on top of the physical robot. The monitor offers a view into the real world from the perspective of the robot's position. This way of displaying the robot's view of the world is an effective mapping in that it is often reasonably and quickly clear to the operator what is being displayed. For example, the source point of the view and the image content are clear from the spatialized context in which they are presented. The dynamic behavior of the monitor is also clear. When the robot moves, the virtual robot agent configuration is updated and thus the monitor containing the video is moved as well. Since the video camera moves with the physical robot the video stream on the monitor is also continuously updated as the robot moves. This is a consequence of the coupling between the robot's physical movements and those of the virtual robot agent in the virtual environment.

The operator has the choice to lock her avatar movements to the robot's movements or to roam freely about the environment. When the user's view is locked to the robot's movements it is as if the operator uses the robot as a

[1] This work has been done by Olov Ståhl.

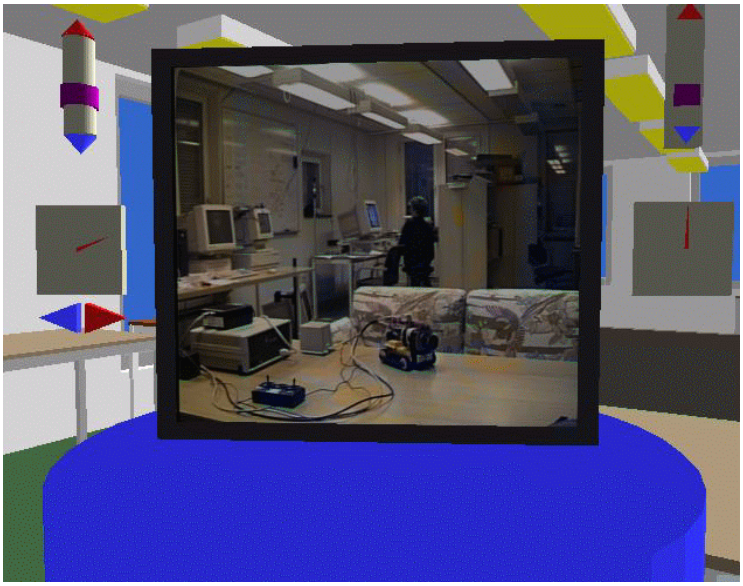

Figure 4.7: The figure shows the monitor metaphor with the live video stream of the real world. A number of robot interaction tools can also be seen on either side of the monitor (see also color figure **A**). *(image: Frécon)*

virtual vehicle to move around both the real and virtual environment. Here, the monitor is fixed to the operator's viewpoint, offering a view into the robot workspace. As the robot/operator pair rotate and translate around the room, the operator is given an interactive video scan of the remote environment. This can work to increase the impression of virtual presence. However once the robot turns away from a scene, the video images of that particular view are lost. There is then a need to leave video images in the environment so they can be later re-visited and re-examined. This need leads to the next solution of using projection screens.

Projection screen

A large surface with a video still image from the real world placed in the virtual environment is another way to place the video within the virtual environment. We call such a surface a *projection screen*. The projection screen can be updated from the camera. Thus the projection screen becomes another monitoring source but not attached to the robot as the monitor metaphor described previously. Image stills taken by the robot are left in the virtual environment in a position that covers the area they depict. Alternatively, if the video source is from a secondary camera, the video stream could be incorporated within the

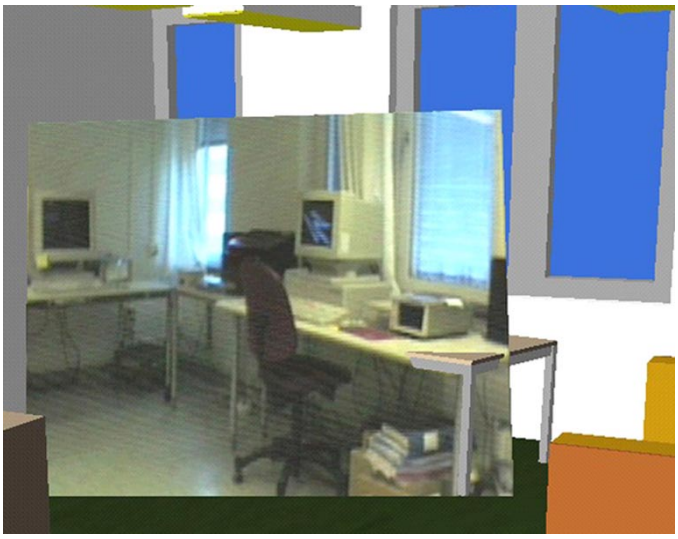

Figure 4.8: Projection screen. A large surface with a video still image of the physical environment is introduced into the virtual environment to display an image of the physical space. Note that the virtual table intersects with the real one. Since the image is taken from a particular viewpoint it is only in correct perspective from the equivalent viewpoint in the virtual environment.

virtual environment and viewed on the projection screen.

One key difference between the projection screen and the monitor metaphor is that the screen is statically placed within the virtual environment and does not move around as the monitor on top of the robot does. In this way, images previously taken can be stored spatially for later viewing. This is a function that the monitor metaphor did not readily afford, a clear and easily understood method for revisiting previous frames. As opposed to a method that enabled a user to "flip" through saved images, the projection screen preserves the spatiality of the CVE interface.

The projection screens also affords a large viewing area within the virtual environment, offering the user an opportunity to view the real world content from a greater distance then the Monitor Metaphor would allow. Typically, the projection screens are "wall-sized." They can be compared to standard texturing techniques where the key difference is that a projection screen is often a 2D object in 3D space within the virtual environment but displays 3D scene structure. Texturing techniques usually place a 2D image onto a planar surface or curve in the 3D world.

There is no obvious way to show the connection between the projection screen and the virtual objects whose physical counterpart is mirrored but cov-

ered up by the projection screen. One way is to allow the the projection screen to intersect with the virtual objects (see figure 4.8). Such a positioning offers the user a visual clue to the connection between the virtual object and the real counterpart. However there are limitations related to the projection screen's 2D nature and the 3D spatial nature of the CVE.

Limitations

There are several limitations to the methods described above, related both to space and time. For the monitor metaphor there is a limitation in the view angle of what might be potentially viewable in the remote space. To obtain an accurate visual image of the remote environment the operator would have to command the robot to pan around the room, even if the robot's camera had panned that space before. Thus, this spatial limitation also relates to a temporal limitation. To pan that space takes real time as the robot has to physically move. In addition there is no clear way within the metaphor to view a history of images already seen. The monitor metaphor also restricts the user to be situated close to the robot if he/she does not want to loose the connection to the real world. Therefore she cannot change her avatar's position in the virtual environment and still be able to see the video in the real scene without moving the robot.

With the projection screen, textures placed into the scene with a fixed flat view of the remote scene from a particular instant in time and viewing angle. There are two major limitations with the projection screen. First, it is only perspectively correct from the virtual environment location corresponding to the point, or near the point, where the image was taken[2]. At other positions, the illusion of three-dimensionality is broken and the connection between the 3D model and the image is not evident. Secondly, with projection screens it is hard to make the associations explicit between the objects on the screen and the objects within the virtual environment.

Over time, there is no image history and, in addition, it takes time to re-acquire images of places previously visited even if those scenes remain unchanged. Though we could save video segments or images for recall, it is not clear how to present these images to the user. They could again be displayed through the monitor interface via a button-operated interface. In that way the user could flip through those images for reference. The problem of storing images was partially solved by using projection screens, but the scene quickly becomes cluttered. In the MarsMap system they attempt to remedy this cluttering by turning on and off the projection screens (called *billboards* in that system) [20]. However that "kludge" remedy does not seem adequate. In short, these solutions do not exploit the 3D nature of the virtual environment. These limitations are what led to the concept of Reality Portals that offers

[2]Note that this only holds if the camera and viewpoint characteristics, *e.g.* extrinsic and intrinsic parameters, are the same or similar.

Figure 4.9: In this figure the Reality Portal method for visualization is shown. The position of the camera and robot are synchronized to the virtual environment and parts of the camera image are extracted and placed in their appropriate places in the world model. In this image there are RP textures on the windows, the computer screen and the white-board on the right wall (see also color figure **B**). *(image: Åkesson)*

solutions to these problems by making direct use of the 3D spatial nature of the CVE.

4.3.2 Reality Portals

The limitations suggested above pointed toward an implementation of another more general solution. This solution is based on applying the appropriate segments of video onto the actual corresponding virtual objects in the virtual model. We call these video-augmented objects Reality Portals, as they are viewpoints or portals from the virtual world into the real world.

By using textures created from images of the actual physical objects the user is offered richer detail and may not encounter the same problems in making the associations from the real to the virtual world. The three-dimensionality of the virtual environment is used in a more sophisticated way as the whole virtual room is used to visualize the video. By laying textures in space it is also possible to have an image history located around the room. If the Reality Portals can supplant some of the uses of the monitor metaphor there is also an added benefit in a great reduction in the amount of data distributed between multi-user nodes of the CVEs. This is especially true if the environment is not especially dynamic.

To demonstrate this technique, a Reality Portals prototype has been developed. The prototype can, with proper mathematical camera model and a reasonable model of the environment, apply pieces of the extracted video im-

ages to corresponding objects in the virtual environment. This works by first specifying special flat objects in the virtual model, the actual Reality Portals. The camera on the robot is augmented with a view-cone that represents the viewing angle of the robot. When the view-cone of the camera intersects one of the reality portals, an event is generated. This event can trigger the request by the reality portal for the appropriate piece of the video image in the video stream. This requested piece is extruded from the stream, cropped and formatted for use as a texture. This texture is then applied to the requesting Reality Portal.

Through this process, textures are applied automatically in near real-time in the virtual world. It is "near real-time" since the prototype currently only manages to generate 1-3 frames per second. However if attention were paid a faster version could be made because there is much room for optimization. As the robot explores the world, these textures are automatically added to the virtual objects and stored in the virtual world database. Thus the time-history limitations mentioned before are partly solved in that old images are placed in the virtual space in their corresponding positions. The virtual world model offers an intuitive spatial sense of how to view these video images and their source. It is much like having a 3D video to navigate around (figure 4.9).

Some of the space limitations are also solved by the process because now the operator is not limited to looking at the monitor at the video of the remote scene, but can instead navigate through the virtual world and see the video spatially displayed. Because highly structured images are split up, many of the video images applied on the Reality Portal are portraits of flat surfaces. Thus there are fewer problems in losing the illusion of three-dimensionality than there were in the projection screen solution.

Note that the textures are extracted from the video image and applied only to the requesting surface. For example, to cover an entire 3D object in the virtual environment, *e.g.* a cube on a table, Reality Portals would be placed on the five potentially visible sides. As the robot navigates around the space and the camera's view-cone intersects with those real surfaces, the textures will be extracted and laid onto the virtual thus object covering it with textures from the video of the real space.

These different forms of using video in the virtual environment can also be used together. The operator can view the real-time video from the monitor interface while also viewing the history of video images via the textures in the virtual environment.

Creating Reality Portals

The method to extract the textures for the Reality Portals from video is based on a basic camera model, a virtual model of the real scene and basic image processing. The camera model together with the virtual model make it possible to predict where in the video image different objects appear. The database,

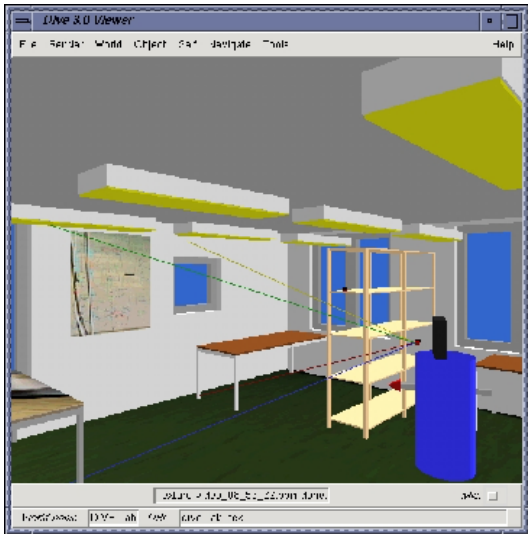

Figure 4.10: This is a visualization of the the camera's field of view on the scene. The rays cast out from the camera on the robot show the limitations of the field of view. On the wall was a Reality Portal that was waiting for the camera to pass over and send it the image of the whiteboard. *(image: Åkesson)*

which stores the definition of the virtual environment, provides the coordinates of surfaces within the virtual environment. These coordinates are transformed through the mathematical camera model, which gives the coordinates of the virtual surfaces in the image. Through image processing it is then possible to extract textures from those areas in the video image. This technique to predict objects within the video image is rather similar to the ones used in Augmented Reality.

Camera Calibration

Every image taken of a scene through a camera is a distorted version of the real scene. With a mathematical camera model it is possible to describe these distortions to some approximation. The most obvious effect comes from the perspective projection, the 3D to 2D transformation. There are also distortions from the camera lens, the CCD array and the video frame-grabber (these are the most significant factors). Point projection is the fundamental model for the perspective transformation wrought by imaging systems such as our eye, or a camera and numerous other devices. To a first order approximation, these systems behave as a pin hole camera, *i.e.* the scene is projected through one single point onto an image plane (the same model used for 3D image rendering). The camera model employed is based on the pinhole camera model but also

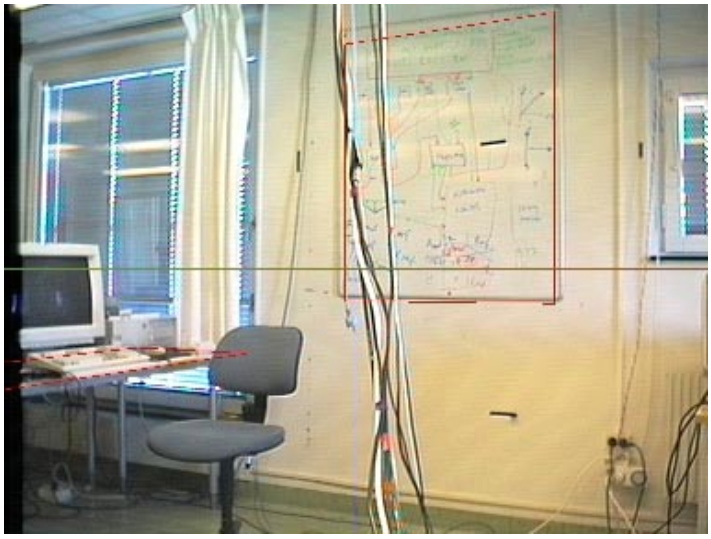

Figure 4.11: The image shows the same scene as in the previous figure, only taken directly from the camera on top of the robot. The highlighted area is the part of the image to be extracted. *(image: Åkesson)*

takes care of some lens properties and effects from the frame grabber.

The parameters in the camera model are not readily available but can be discovered through a camera calibration process [113]. The process of camera calibration has been studied intensively during the last decades in both the photogrammetry and computer vision communities. The parameters to be discovered can be divided into two different sorts. These are extrinsic and intrinsic parameters. The extrinsic parameters belong to the setup of the camera, *e.g.* the estimation of camera position, rotation and translation. Extrinsic parameters represent the relationship between the coordinate system of the camera and the global coordinate system. The intrinsic parameters include the optical and electronic properties of the camera such as focal length, principal point, lens distortion, the aspect ratio of the pixel array and other CCD effects.

The extrinsic and intrinsic parameters must be known in order to use the camera model and predict where 3D-coordinates will be mapped onto the 2D-image plane. To do real time measurements of the parameters is almost impossible, as the calibration process is computationally demanding. If the intrinsic parameters do not change during and between runs it is enough to calibrate them only once. By turning off features like auto white-balance and auto-gain and locking lens properties such as zoom and focus, the electrical and optical properties of the camera will mostly not alter during or between each run. Correspondingly, neither will the intrinsic parameters change and it is enough to

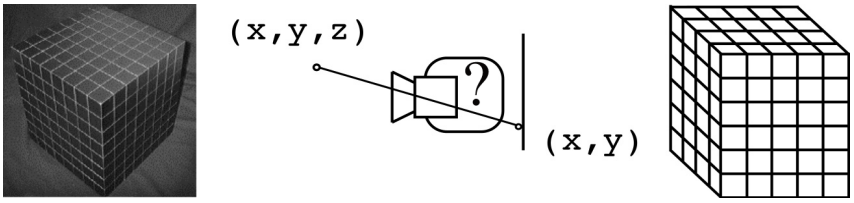

Figure 4.12: Camera Calibration. The extrinsic and intrinsic parameters of the camera model need to be discovered through a calibration process. This process is facilitated by capturing an image of a precisely known object. The camera model is then used to transform the 3D coordinates of the virtual world into 2D image coordinates. The method used was provided by M. Li [69]
.

run the calibration process once.

For a camera which does not move, the extrinsic parameters can also be calibrated once. If however the camera moves it is necessary to update these parameters, which is difficult to perform in real-time through camera calibration. Though rotation and position are easily measured units from our robot, due to drift and sensor inaccuracies we can only get accurate real-time measurements of relative changes from a known start position for a limited amount of time. We add the robot starting configuration (coordinates plus orientation and height) to the camera start position discovered through the calibration process. Thus we have start values for all the parameters needed for the camera model.

The actual calibration process is begun by capturing an image of a scene with known 3D coordinates. The parameters are then calculated based on a comparison of where these coordinates appear in the image and where they should have been. In our case we use a cube with a grid of vertices painted on the surfaces and through edge detection these are found and the parameters can be calculated (see figure 4.12). Unfortunately due to the reasons given above, the registration of the image and the virtual world degrades over time, especially for images taken from a distance where the a pixel represents more linear space. Thus either a certain context would have to be defined or a different dynamic on-board calibration process would need to be discovered. Details on experiments with calibration can be found in Karl-Peter Åkesson's master's thesis [5]. The right solution would involve a dynamic self-localization process based on information in the CVE model that in turn defines, or progressively refines, the camera's extrinsic parameters. Making such a system robust is left for future work.

Figure 4.13: Texture Extraction. The left image highlights the Reality Portal areas to be extracted (table and whiteboard). In the right image, the whiteboard image has been extracted and placed as a texture in the virtual environment. *(image: Åkesson)*

Texture generation

The actual process to generate the textures for a Reality Portal is based on standard image processing and transformation algorithms. In the virtual model the 3D-coordinates of different objects and definition of object surfaces within the virtual world are given.

As the virtual model is a representation of the real environment, these virtual surfaces are the same as the ones for the real objects. The camera model, with calibrated camera parameters and continuous updates of the extrinsic parameters, is used to do a standard transformation from the 3D-coordinates in the virtual world to 2D-coordinates in the video image plane.

The parts of the video image which belong to different surfaces can now be predicted in the 3D virtual environment. Depending on the graphics system used to render the virtual world, different methods are employed to extract the texture. The most common graphics system, OpenGL, only supports rectangular images as textures and therefore a non-rectangular texture for the Reality Portal has to be re-sampled using bilinear interpolation (a textbook image processing algorithm). This sampling is used to make a non-rectangular Reality Portal rectangular by adding pixels. Such problems are also encountered when the desired image segment is in perspective in the video image. Such a segment will be extracted as a non-rectangle and needs to be warped to fit as a texture on the Reality Portal. This is the case, for example, when a square object is viewed from a side angle. Note that pixels have to be added, and thus parts of an image texture may have different clarity as a result of the warping.

The first step in the process is to calibrate the camera. The calibration process can be achieved by one of the known processes. Methods employed in

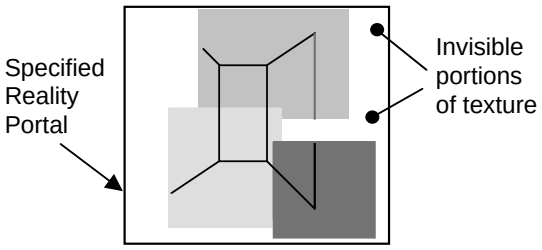

Figure 4.14: Partial Transparency. A specified Reality Portal might contain surfaces not visible in the camera image. Portions outside the known camera view are textured transparent. Stacking several image segments can, over time, complete the Portal.

this work are based on [113, 69]. The robot camera used can be said to be active in the sense that it sits a mobile platform. While navigating around the environment, the camera acquires images of the world and the controlling process is capable of extracting video textures from the camera signal. The image-textures are acquired either by the event of camera movement, user choice or periodically. Any of these events triggers the camera to produce textures for each Reality Portal within its field of view. The camera records its own position and the position of the Reality portals and uses that information to extract data from the camera signal (image stream). The camera process then uses these images as textures. This process is accomplished by using the virtual object geometry coupled with the camera parameters (extrinsic + intrinsic) and projecting the relevant object surface onto a virtual camera image plane matching the real camera. Using this 3D to 2D projection, the relevant portion of the camera image can be extracted as a texture and used in the virtual world. Thus as the camera pans around a room, Reality Portal objects receive textures. As the camera moves around the environment the virtual world fills up with textures and comes to resemble more closely the real world scene. The quality and accuracy of the textures depend on the following factors: calibration, the distance the camera is from the object, lighting and the angle the camera makes with the objects surface.

Partial Transparency

A possible problem with extracted textures is if the entire Reality Portal surface is not visible from a particular camera view point. If there are segments of a Reality Portal that lie outside the field of view of the camera, the texture will be transparent. In the event a whole surface of an object is not seen from a camera, the remaining part of the Reality Portal will be transparent. Thus, a layer of Reality Portals with textures from different cameras can be placed on the surface of the object. In this way, segments of Reality Portals that are

not seen by one camera position will show underlying Reality Portals textures and the result will be a an overlapping surface composed of a number of video textures.

The Reality Portals extract the textures from one camera. If several cameras were used, or if an image is taken from a different viewpoint, more then one camera image is taken of the physical object. It is not obvious how the texture for the Reality Portal should then be generated, *e.g.* from which source? One solution is to use image processing to mix the parts from each camera image into one texture. Partial Transparency is a method we have explored that gives a decent result and which is much more efficient as it does not use any image processing to achieve the result. Each Reality Portal is associated with only one camera so it will only receive textures from that camera. Parts of the texture for an object that are outside the view of the camera are made transparent.

By stacking Reality Portal objects on top of each other it is possible to generate an object that has a texture with an image of the whole physical object even if it is not completely imaged by any camera. Parts of the texture that are transparent, *i.e.* that are not imaged by a particular camera, allow textures from Reality Portal objects beneath, to be viewed.

Another key component of the system is warping the texture onto the object. Because of, for example, camera skew, an extracted camera image segment and texture may not contain right angles. Such images will first go through a warping stage to fit onto the Reality portals rectangular surface.

Freezing the Reality Portal in time In the current version of the prototype it is possible for the user to freeze each Reality Portal, *i.e.* instruct it not to take any new textures. In this way it is possible for the operator to save snapshots of the environment. The operator can look back to previously visited areas without needing to steer the robot to that specific location or pan the environment with the camera. This freezing may be desired if the user wants to preserve some information and not have it overwritten the next time the camera's view-cone passes over that area. In the current prototype the operator clicks with the mouse on each Reality Portal in order to freeze it.

4.4 Command Interaction

The presentation of the interface in this paper centers on visualization and interaction. This interaction occurs between the robot and the operator by way of the virtual environment. Thus there is interaction between the operator and the virtual environment and interaction between the virtual environment and the robot. In some instances this interaction medium is made transparent, to encourage the operator to concentrate on interacting with the robot, at other times the interaction between the robot and the virtual environment is

more explicit. In the former case, these are when the human is controlling and commanding the robot to move to certain locations (*e.g.* point-to-point control) in the latter case, this is when the human is interacting with the robot to produce finer control of robot movements. Another situation is when this interaction happens automatically between the virtual environment and the robot. This automatic interaction happens by way of spatial embedding of routines within the virtual environment and this spatial embedding, or spatial programming is enabled by the concept of the *spatial model*. Finally, similar interactive functionality is enabled by the speech interface to the robot.

These different methods of interaction, the direct navigational commands by the robot, the spatial programming and the speech interaction are discussed in turn in this section.

4.4.1 Direct and deictic interaction

The least abstract form of interacting with robot is the interaction that takes place through pointing and clicking in the virtual world. These are represented in the CVE as 3D analogies of traditional 2D WIMP (windows, icons, mouse, pointer) desktop *direct manipulation* concepts common on 2D desktop platforms [96]. For previous work and a discussion of system that implements a 2D shared desktop in a shared 3D system the reader is referred to [38].

There are primarily two different types of interaction mechanisms that implement a "point and click" style of interaction in this system. These mechanisms differ primarily in their level of robot control. The first, at a low-level of robot control, is an interface to control the robot to take the basic movement actions of "forward," "backward," "rotate right," "rotate left" and "stop." This interaction happens by way of the arrow interface depicted on the left side of the image in figure 4.15. The other mechanism for specifying navigational points is that whereby the user clicks on a point on the floor, commanding the robot to move to that location.

Direct interaction The first mechanism is depicted as graphical icons that represent and implement the functionality of basic atomic movements of the robot. These movements mirror the basic fundamental movements of the physical robot platform (forward, backward, rotate right, rotate left). Each of those four directional command types are represented graphically as arrows that point. These follow the street navigational conventions where an "up" arrow signifies forward, left and right signify left and right turns and a "down" arrow signifies backward. In the CVE this object is tilted slightly forward off the vertical plane to indicate this forward/backward motion a bit more. In the 3D graphical world the limitation for this depiction is much the same as in physical traffic navigation world. An arrow pointing forward, or backward, in the horizontal plane would convey little information to and be problematic to interact with for a user that is oriented upright in the Y-plane. Thus a

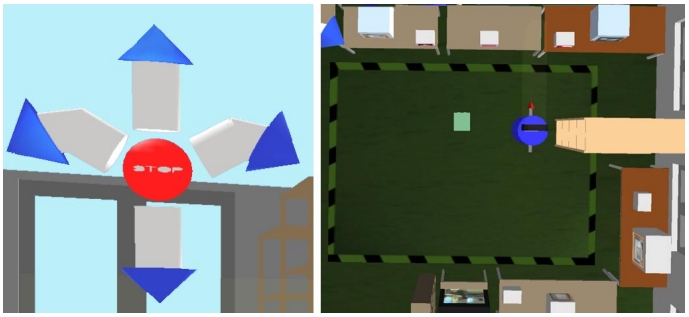

Figure 4.15: The image shows the two types of direct manipulation mechanisms in the system. On the left is the system for controlling the robot a the level of issuing basic atomic navigational commands (forward, backward, rotate-right, rotate-left). On the right is the more abstract mechanisms of pointing on the floor to specify a point-to-point navigation command.

similar convention to that given in everyday navigation of the physical world is employed in the virtual world and the control mechanism is oriented nearly upright. The user commands the robot by clicking on an arrow which results in a command being sent via the virtual robot agent to the physical robot. The physical robot's communication software receives this command and moves selected direction by a small increment. These increments are built up (and produce fluid motion) by either repeated clicking or the holding down of the mouse on the arrow.

Deictic interaction The second mechanism, in contrast, is both higher level and more transparent to the user. It is higher-level in that it implements a higher-level of functionality in the robot, that of point-to-point navigation. It is more transparent in that, unlike the arrows, the functionality is not made graphically explicit to the user. Instead the functionality is embedding in pre-existing objects that, when the user is informed or discovers the functionality, afford interaction in the intended manner. In this case it is the graphical object of the floor that serves a number or purposes in the user interaction which is a representation of the physical object of the lab floor. The floor provides the medium through which the specification of navigational horizontal planar goal-points are specified. The user commits this command by clicking on the floor which results in a coordinate in the horizontal plane being sent to the robot as a new goal for navigation. The points on the floor then serve as the deictic reference for navigation.

Both of these mechanisms work by spatial properties of the 3D objects. The object that intersects this ray first, relative the screen, becomes the selected object. In the case of these particular mechanisms, it is the co-location of the

mouse pointer (also a 3D object) and the intended control objects (*e.g.* the directional arrows, or the floor). This co-location occurs by casting an orthogonal ray from the mouse pointer, which sits on the 2D surface of the screen, into the virtual world. This co-location accompanied by a mouse click triggers an event that then triggers a set of functions that are implemented in that object.

Reviewing control modes

Steering the robot with direct control: Using the mouse interface the operator can open a control panel that allows the operator to steer the robot via directional arrows. This is essentially a remote direct control interface and very limited. A better solution to manual positioning is to use the 3D context to push the robot around the ground plane and position it on the floor. This can be achieved by different methods. One is by exerting forces on the robot during its real-time motion through the environment, another is to specify a desired configuration (X,Y,Θ) by placing a "ghost" model of the robot in the desired final position. This latter positioning can, of course be viewed from any angle including top (plan) view of the virtual space.

Pointing in the fishbowl: In this mode the operator points at objects in the virtual world as positioning goals for robot navigation. In this way also the operator could specify a path by a series of navigational landmarks. This method allows increased robot autonomy in a way that the manual manipulation techniques do not. As the navigational competence of robot increases the operator can specify goals closer to the task level, *e.g.* final configurations where the path-planning is left to the robot. That is to say that the pointing metaphor is quite extensible with respect to increased robot autonomy.

Pointing in the real world: This has not yet been implemented but the essential framework exists from the Reality Portal setup to do simple interaction in the video images. The model of the world is calibrated to the camera position. This in effect gives a way of finding 3D spatial points in the model if image coordinates are given. In this mode the operator would point at objects in the real world (specifying a 2D image coordinate) and rely on third-eye or robot camera calibration to give the world position. By pointing at objects in the real world, the location of the gesture in image coordinates can be acquired and transformed into a 3D coordinate by ray casting into the VR database from the camera's origin.

Manipulating a graphical object in video: Using the same RP framework as described above, with an alignment between the real and virtual worlds the operator can also manipulate a graphical object in the video image to specify the final position for an object. In figure 4.16 this is being done with

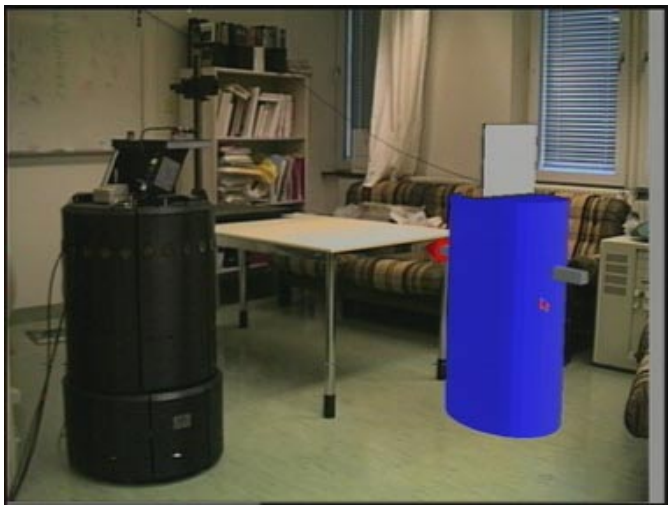

Figure 4.16: The image shows a graphical robot being positioned in the real world via augmented reality techniques. Given position calibration between the real and virtual worlds the position of the graphical robot can be used to give the robot update pose (X,Y,Θ) instructions.

a graphical representation of the robot to specify the desired position of the robot.

4.4.2 Spatial model

Inside the virtual environment that the DIVE system implements, there is a notion of spatial interaction [17, 16]. This model provides a method of interaction for the operator, the robot, and the objects within the virtual and real worlds. In this section this spatial interaction model and the methods it suggests are described.

Here the key concepts are summarized which constitute the DIVE spatial model of interaction, the details for this model can be found in [16]. The goal of the spatial model is to provide a small but powerful set of mechanisms for supporting the negotiation of interaction across shared virtual space. The spatial model, as its name suggests, uses the properties of space as the basis for mediating interaction.

The first problem in any large-scale environment is determining which objects are capable of interacting with which others at a given time. **Aura** is defined to be a sub-space which effectively bounds the presence of an object within a given medium and which acts as an enabler of potential interaction. Objects carry their auras with them when they move through space and when

two auras collide, interaction between the objects in the medium becomes a possibility. It is the surrounding environment that monitors for aura collisions between objects. When such collisions occur, events are triggered that allow listening programs to take the necessary steps.

Once aura has been used to determine the potential for object interactions, the objects themselves are subsequently responsible for controlling these interactions. This is achieved on the basis of quantifiable levels of **awareness** between them. Awareness between objects in a given medium is manipulated via **focus** and **nimbus**, further subspaces within which an object chooses to direct either its presence or its attention. More specifically, if you are an object in space the following examples help define the concept:

Focus The more another object is within your focus, the more aware you are of it;

Nimbus The more another object is within your nimbus, the more aware it is of you.

The notion of spatial focus as a way of directing attention and hence filtering information is intuitively familiar from our everyday experience (*e.g.* the concept of a visual focus). The notion of nimbus requires a little more explanation. In general terms, a nimbus is a sub-space in which an object makes some aspect of itself available to others. This could be its presence, identity, activity or some combination of these. Nimbus allows objects to try to influence others *i.e.* to be heard or seen. Nimbus is the necessary converse of focus required to achieve a power balance in interaction. These concepts of focus and nimbus form a basis for the programming of spatial functionality in the CVE.

Spatial Programming

This spatial model is used to create an interactive and informational rich immersive environment that stores the methods to aid the robot's interaction in the real world. The concepts of aura, nimbus, focus are key to the way the robot interacts with the virtual and real worlds. Using the concepts of spatial boundaries and auras we can define interaction mechanisms and methods for sharing information between the robot and the environment.

A concept of *spatial programming* is introduced that is similar to the concept of *Object-Oriented* programming. It is similar in that routines are encapsulated as objects, divided up and distributed into routine packages where they can be called to produce their output as way of achieving modularity in programming. However instead of messages being sent to the objects to initiate programmed functions, the object routines are triggered by spatial events. A *spatial event* is a signal that is generated when the proximity or position of an object meets a specified condition. An example of such an spatial event is the generation of

a signal when certain objects reach a specified proximity to one another. In all the events in the current system such an event is triggered by the collision of two objects. Proximity of such objects are achieved by using the aura, nimbus and focus aspects of the spatial model. Thus an object may specify a nimbus and focus, which are virtual object a defined volume, that when a state of co-proximity is reached, an event is triggered. These objects may be rendered or may become visible depending on context.

In conjunction with an event, pre-specified routines are executed. Thus the concept of spatial programming entails the specification of spatial model objects, the specification of events based on certain spatial conditions, and the specification of functional routines to be executed when the conditions are met and the event occurs. Thus movement in space is analogous to messages sent to objects in object-oriented programming with the result that a certain routine is executed.

The difference in this programming mode from that of traditional models is that routines are localized to those objects within the routine's scope. Also that, from the programmer's standpoint, the programming task can be parceled up into programming tasks where the division is explicitly spatial. The belief is that there is a benefit in this division, and that like object-oriented programming, the program can find natural divisions of functionality that allow the details to be encapsulated into well understood and higher-level divisions of functionality. Also for the returning programmer the topography of routines is more easily assimilated and understood, because the routines deal in functions where their context is spatial. The spatial positioning can enforce the understanding of the spatial function.

For example using the concept of object aura a means of transferring information for navigation and object identification can be defined. If the robot's aura collides with an object's aura that object may then open up a channel, *i.e.* the robot focuses and the object projects nimbus, thus enabling the object to pass information to the robot that would be pertinent to the mutual interaction. In this way each object stores information and methods about itself. This information can include: object identification; object function; navigational approach method; grasping method; and recognition method. These last three types of information deserve special mention. An object may store the actual methods in which to perform a local interaction such as recognition. Given that the position of the object and the position of the robot are well known these methods can be rather specific.

Likewise, using the boundaries in space, various locations in the environment may store information and methods regarding navigation. For example there may be certain areas of the environment where great care must be taken, so crossing a boundary could then act like entering a "speed control zone" and thus negotiate control for the robot's velocity. Similarly there could also be areas in the environment where certain configurations or specific paths should be avoided or taken. Crossing a boundary into such an area would open up a

channel to transfer specific navigational commands to the robot.

Using this model of interaction reduces the complexity of the robot control process, *e.g.* from the need to have knowledge about the nature of specific locations in the environment. Using spatial programming we are distributing the processes and specific information throughout the environment. Also using the model in this way it makes it much less necessary for a robot to store knowledge about a new environment before actually entering it. Thus when the robot crosses the boundary into a new environment it or the user would be given all the necessary global information regarding that world. So while distributing the knowledge is good from reducing the apparent complexity of a program, it also beneficial in unloading the functionality of the robot agent itself. In this way the robot can be seen bumbling through a world of smart objects that may know how to take care of it while it moves spatially. The model of the world, in addition to embedding the programmed functionality, also represents to the user the robot "knowledge" of the world. That is what objects and functions the robot has access to or that will influence the robot's behavior.

The direct manipulation navigational interface mechanisms can also be seen as implementing a form of spatial programming. Each of those interface objects (arrows, floor) are in fact 3D objects in the model description language. In the 3D world and model of events, the co-location and mouse-click causes a *selection* (as opposed to collision) event to occur which then triggers specified functions which are embedded in that particular object which received the selection event. Inside the description of that object is contained programming code (in the TCL language) that then executes communication with the robot through a socket and TCL-C interface.

4.4.3 Speech interface

Hitherto, controlling and manipulating a virtual or augmented reality has mainly been through *direct manipulation*, an interaction paradigm based on immediacy of control and tightly linked feedback between action and effect. According to Ben Shneiderman, direct manipulation interfaces generally share three main characteristics: 1. continuous representation of the object of interest, 2. physical actions or labeled button presses instead of complex syntax, and 3. rapid incremental reversible operations, whose impact on the object of interest is immediately visible [94].

These characteristics have usually been seen as standing in contrast to command based interfaces that build on more expressive forms of input such as formal command languages or human languages. While Shneiderman's points certainly have been understood as a justification for a completely analog interface representation such as pictures and graphs, and analog operations such as gestures, points one and three do not in fact in any way contradict the possibility of using text or other language input – indeed, any interface at all, be

"go forward until I say stop" This command instructs the robot to move forward and listen for the signal to stop.

"come to me" This command asks the robot to locate my avatar and move toward me in the virtual environment.

"move here" This command asks the robot to come to my virtual position in the virtual environment.

"move there" This command asks the robot to move to the place gesticulated via the user's mouse.

"go to the table" This command asks the robot to move to an object in the virtual environment.

Table 4.1: Description of a hierarchy of deictic commands to the robot.

it language based or point-and-click based would do well to follow the principles. We will use language, in our case speech or typewritten text, as one of the mechanisms of interaction, thus relaxing the constraints posed by Shneiderman's second point, but continuing to observe points one and three. Language, as shown below, is necessary to manage the level of complexity following from instructing a robot.

The spatiality of virtual environments offers the operator intuitively useful means of selecting and manipulating objects in the vicinity, much as gestures do in real life. Deictic reference, such as "this" and "that" are easily defined and formalized in virtual reality. They allow the operator to refer to entities other than concrete objects, *e.g.* concepts, past and future events, actions, etc. For more detail on this discussion and motivation the reader is referred to our joint paper [103].

The motive for including language in a robot-control interface is to add a level of abstraction to the system: to be able to specify goals on a higher-level than pointing at visible objects. This, of course, presupposes a level of representation abstract enough for symbolic reasoning; this is achieved through the explicit model of the robot's real world knowledge in the virtual world.

Examples of commands that have been speech implemented are "come to me," "go there," "rotate until I say stop.". These are each difficult to express by mouse gestures alone. The next step is to work with commands such as "Go to the table," and compound commands such as "pick this up and bring it there," which encompass navigational commands with a grasping and releasing command.

Why Charades Are Difficult: As given in Ivan Bretan's work and collaborative work we have done together [26, 60, 103], there are cases where speech is

clearly more fit for the specification of certain conditions. The classic example of this is the following: in a space of numerous differently colored marbles, the user may want to make the simple request "Bring me the red ones." In such an instance, direct manipulation by selecting each red marble separately burdensome. In such cases, speech when available, can be seen to be more appropriate. Thus motive for including language in a robot-control interface is to add a level of abstraction to the system: to be able to specify goals on a higher-level than pointing at visible objects. This, of course, presupposes a level of representation abstract enough for symbolic reasoning: we have achieved this through the explicit model of the robot's environment in the virtual environment. The virtual environment offers a shared representation through which the robot and the user can interact and share deictic references.

The implemented speech sub-system: DIVERSE (DIVE Robust Speech Enhancement) is a speech interface to the virtual reality platform DIVE. DIVERSE is developed at SICS for use as a test system to experiment with multimodal interaction [60]. DIVERSE allows for spoken language control of operations that are normally carried out through direct manipulation in DIVE, such as transportation of objects, change of view, object creation, deletion, coloring etc, while still retaining the possibility to perform actions through direct manipulation whenever that is more suitable [26].

Interaction in DIVERSE is mediated through an animated agent to allow explicit modeling of the linguistic competence of the system, both in terms of output language and in terms of the gestures.

DIVERSE was expanded and re-written in OZ by Thomas Axling and Scott McGlashan connected to the Nuance speech system and integrated to DIVE through the ODI interface. It was in this system that the speech commands given above were implemented. Unfortunately this system was not maintained and fell out of date and out of license and was unavailable at the time of the Study of Use described in the next chapter.

4.5 Demonstrations and Study of Use

The system has been in use and development since 1995. In that time there have been several formal and *ad hoc* demonstrations. The process of these demonstrations has informed the development both on design and robustness issues. There is however also a need to augment these demonstrations with a more structured look at the system and the relationship users have with it in order to improve the system and gain a better understanding of its use. This section first presents a survey of these demonstrations and then presents a Study of Use that was executed in December 1999.

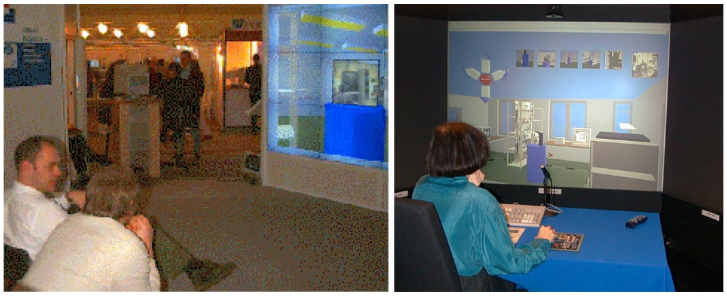

Figure 4.17: Robot System Demonstrations. At CeBit'96 the robot was in Stockholm and could be controlled from Hannover (left). A study participant interacting in the SICS Grotto with the remote robot system in December 1999 (right).

4.5.1 Cebit 1996

In March 1996, the system was demonstrated at the CeBit'96 international electronics fair in Hannover Germany. The demonstration was conducted for a few hours a day for a number of days in the GMD booth of the research display hall. The CVE with the model of the robot working environment was displayed on a video wall and the human operator was at CeBit and the robot was located in Sweden. In this demonstration the robot could be moved around the remote space and its position visualized in the CVE. Through the remote visualization, via the monitor metaphor, contact was made between CeBit visitors and persons situated at the robot lab in Stockholm.

4.5.2 Ad hoc demonstrations

Over the years a number of informal demonstrations have been made in the ICE lab and in the Grotto multi-screen display at SICS. These demonstrations have been for groups such as institutional steering groups at SICS, visiting researchers, industrial representatives and other external research parties. Such demonstrations have varied in length from short presentations to longer term functioning displays as part of local events such as "Industry Day" at SICS.

4.5.3 Study of Use

One purpose of this study is to address a lack of robot studies that seek to gain an understanding of the relationship between how people use robot systems. Specifically how they bring their expectations and everyday methods to bear in using a system in performing a task in collaboration with the machine. Many studies exist that seek to measure psycho-physical parameters and task efficiency. Such studies often answer different research questions than the ones

put forward here. Here the *Study of Use* emphasizes the use of the system for a particular task between the human supervisor and the robot assistant. It is primarily a design study aimed at improving this system, identifying issues and informing the design of others similar systems.

4.6 Summary

This chapter has outlined the technical components that make up the robot system. The next chapter presents an explorative study of the use of the robot.

Chapter 5

Study of Use: Human-Robot system

5.1 Finding Remote Flags: a Study of Use

To gain a better understanding of human-robot collaboration and to improve the system a Study of Use was conducted on the mobile robot system described in this thesis. This chapter describes first the motivations, then the study configuration and finally explores the data collected. The task was a remote searching task employing the teleoperated robot. The results of the study are primarily based on interviews with the study participants.

There are multiple purposes in performing a study. A primary purpose is to explore the design issues that have guided this work: to discover what works, what needs refinement, what may be discarded and what might be added. In addition to concrete design questions there are other questions about how users approach and relate to this robot system, and where a better understanding of use can help. Then there area number of secondary reasons for performing such a study. Studies raise new questions, heighten and renew enthusiasm for systems, and bring in new robustness issues. For example, performing a study with a working robot that has physical mass brings up new interface issues that may differ significantly from performing a study on a software application or a simulation. Reasons and good arguments for performing such explorative user studies in general, can be found elsewhere (e.g. [80, 81]). The main purposes in performing this study can be summarized as follows:

Design of interface The design of the interface undergoes scrutiny in terms of both validation and in a search for new design ideas. This study's main goals are to discover those aspects of the interface that have resonance with the users as well as to discover new design ideas through discussion with the participants.

Figure 5.1: View of Study participant interacting in the SICS Grotto with the remote robot system in December 1999.

Improve Understanding A question of the study is if there are issues that might be provide a better understanding of this system. It is also the intention that the reader will find aspects of this report that might help in considering the design of new systems of this sort.

Experience of making a "user" system A system used by users for any significant amount of time has different demands on it than a simple prototype. Not only must the system be robust enough to be used for longer periods of time, but it will be subject to less delicate use than would be given by the designer. Also the functions and meanings of the various system features need to be made clear and sensible.

Demonstration of the working system Here the system is demonstrated as a working system, its use explained and its features explored.

This chapter falls into two sections. The first section 5.2 is a description of the study setup and execution in as much detail as is reasonable. This is given in order for the reader to develop a sense for the environment in which the robot was run, who participated and what was recorded. In section 5.3 a discussion centered on system design and user responses is explored. This includes what was learned in the study, and the results of the study. The next chapter provides a more succinct collection of the findings from the study including the implications for design. What seemed to work, what would be done differently and what the next steps might be.

Figure 5.2: The remote robot heading toward the Tibetan flag-box (see also color figure **E**).

5.2 Description of Study

During the third week of December 1999, 14 persons were invited to the SICS Grotto to take part in the Study. Persons were solicited via email announcement and taken, for the most part, on a first come-first filled basis. To work out bugs in both the system and the study procedure, a few pilot runs were made with "friendly" users the week before the scheduled sessions.

The basic format of the study was to have participants work through a remote searching task in collaboration with the remote robot while being videotaped and asked to "talk aloud." The task execution was followed-up by an interview. There was also a written survey given in two parts before and after the task. The study's goals are more grounded in the interest of design than in evaluation or validation.

The participants in the study consist of a number of researchers from various backgrounds as well as non-expert representatives. The purpose is to elicit insights brought about by enrolling participants with different perspectives. The system itself crosses many disciplinary boundaries and can benefit from these different perspectives. Keeping in mind the possible deployment of such systems to the general public, the layman's perspective was also sought. The actual composition included researchers and students of linguistics, cooperative work, HCI, robotics, graphics, computer vision as well as those outside the computer research, e.g. biology and film and the general citizen. It should be noted that while some of the participants were experts in their fields, the non-expert is heard as the expert and not consciously treated differently in the discussion.

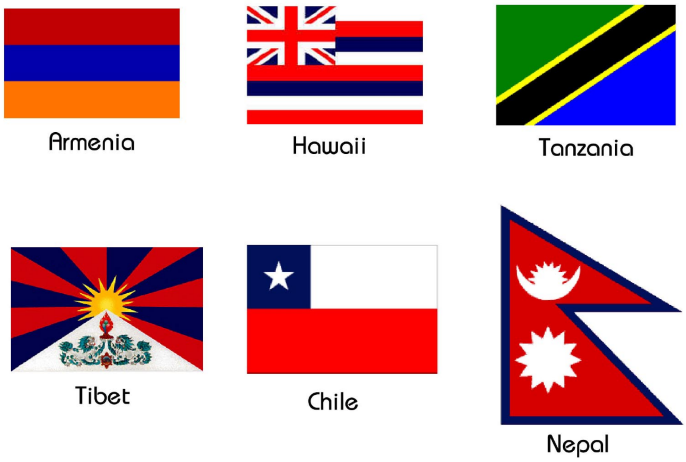

Figure 5.3: Image of information sheet given to users while executing the task. It contained labeled color pictures of the flags that can be found in the real and virtual environments.

5.2.1 The Task

The task was to find flags in a give sequence in a remote space. To do this task the participant is asked to employ the remote robot system in the search. There were six country flags in the remote space and these "flags" were color printouts of country flags pasted onto empty paper supply boxes. The flag-boxes were placed around the room and the participant was to navigate around the space in a particular order. The prescribed ordering was to be discovered. The participant was seated at the desk (figure 5.4 and given a sheet of paper that included the flags in the study (figure 5.3). The participant was told that the first box to find was "Armenia" and that the next box to find would be found on top of the Armenian flag-box (figure 5.2 for an example). The flag-boxes were represented in the virtual world, however the second flag, which signifies where to go next, was not. The participant then had to use the video to realize the ordering. This forced the user to use multiple visualization methods to complete the task.

Explanation of Task To each of the respondents the following text was given orally as instructions.

> You are sitting in the "control room" for a remote robot. The robot is in remote location. In the study, you are asked to imagine that that place is a distant dangerous environment (e.g. with the potential for radioactivity) - a place you would not be able to

Figure 5.4: Image of the desktop the respondents were presented with.

> go otherwise. Your task is to use the supplied interface to navigate around the remote environment and find the flag-boxes in the environment. Each flag-box contains two flags, one large one and one small one. You may find that you have to use different visualization methods to see the flags. Once you find the big flag, the little flag above it indicates which flag you are to find next. The task is completed when you find all the flags, or feel you have "had enough."
>
> During the experiment you are asked to "talk aloud". This means to say what you are thinking, e.g. "I am going use the left arrow to move left so I can find the Nepalese flag, I think I see it in the video."
>
> In this study, criticisms, frustrations and suggestions are just as useful as compliments. The attempt is to understand this system better in order to improve it and other systems like it. So your candid comments are encouraged.

After the introduction text was read the participants were given a brief demonstration of how the system worked, the different ways of navigation, and shown the remote robot on a video screen as it moved. Then the side screens were turned off, leaving only the virtual environment view for the first half of the task execution.

Analysis of Task Admittedly this task could have been done by an automatic system. Although the author does not have experience with such a

system in practice it is believable that such a system could be constructed using state-of-the-art solutions based computer vision and robot navigation. However it is also believed that this study employs one of the more sophisticated systems upon which such a study has been carried out. The intent is that starting with such a task will point the way to further study and research in the context of more sophisticated robot situations employing greater degrees of robot autonomy and competence.

5.2.2 Task interaction

Almost all the participants carried out the task methodically by discovering the flags in sequence. One participant, Magenta, was confused at first and explored the entire space but then understood what was to be done and visited the flags in the prescribed order. Another participant, Yellow, was not able to visit the last flag and complete the task because of a mechanical failure. In the virtual world, all the boxes were initially blank, marking that they were "undiscovered." Because of the start position of the robot, the Chilean and Hawaiian flags were almost always exposed first (the RP system turned them on). These were then often remembered by the participants when it came time to find them. The only flags that presented some difficulty were the Tanzanian and the Nepalese flags. The Tanzanian flag required the robot to cross the room and come around the bookcase. The Nepalese flag was different from the others. The box was in a different orientation (portrait), and was generally not visible from the top view because it was positioned inside a bookcase (as opposed to on a table). This made it harder to find and this was done on purpose. The Nepalese flag-box was the last in the sequence. Sometimes this flag was discovered while searching for other flags. Often this provoked a remark ("ah ha!") because of its unusual position.

5.2.3 Technical setup

The study was executed in the SICS *Grotto*, an enclosed space with three projection screens (figure 5.5). Interaction was done by mouse and keyboard interface. The robot was remotely located in the ICE lab (a room physically 100 meters away). The remote space consisted of more than half the lab with six tables and several bookcases supporting a set of different boxes with different colored country flags imprinted on them. This basic structure was also modeled in the virtual environment model.

All three screens in the Grotto were used. The center screen was the interactive view of the CVE with the virtual robot agent embedded in the model, the right screen was the "third-eye" or camera view, that is a camera observing the lab, where the robot can be seen in context. The third screen was the live video from the on-board robot camera. These side video screens were left off

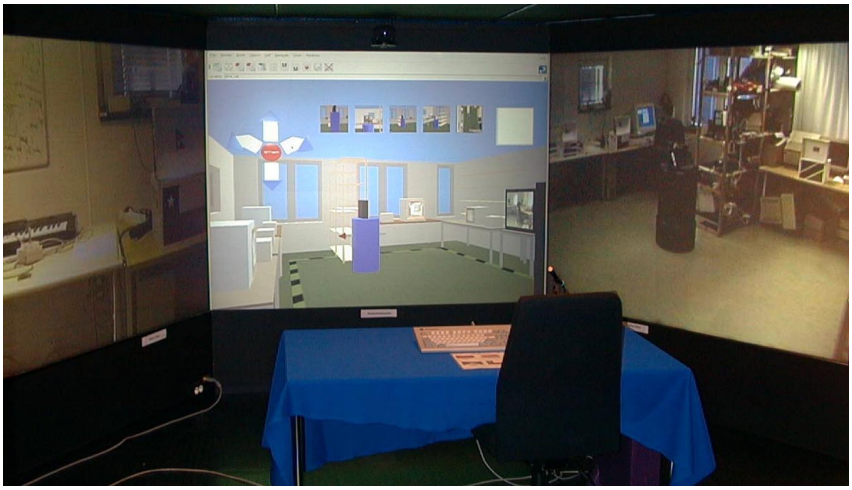

Figure 5.5: A view of the screen layout in the study. The CVE is in the center, the "Robot's view" is on the left and the "Camera view" is on the right (see also color figure **D**).

until approximately half of the search task was completed, this was marked by the discovery of the Tanzania flag-box.

5.2.4 The system functionality in the study

The system used in the study had the basic navigation functions of point-to-point and manual steering. For visualization, video from the remote space was fed from the robot camera and inserted into the virtual world. There were also a number of pre-computed Reality Portals that would turn on when the robot's view cone intersected them. The systems that were not functioning were the arm, the speech interface, and the real-time Reality Portal system. One of the arm motors controllers had failed making it unavailable for the study. The decision not to use the speech system was both from a design and a systems standpoint. Speech was not necessary for this task (explored further in the study report) and the current speech system implementation is not compatible with the current system version. The real-time Reality Portal system is functioning but because of problems of resource contention (this system requires an additional SGI with video input), calibration and overall complexity, it was decided to use pre-computed Reality Portals.

Several times during the study the current session had to be paused so the system could be restarted. This happened more for the first 2 days of the study than for the last 2 days. One of the primary reasons for these breakdowns was due to a newly purchased radio ethernet unit. Unfortunately this caused

new problems do to radio interference, network delay and broadcast traffic, conditions for which the system had not been tested. The new video link and the ethernet used similar frequencies for transmission causing interference. When this was realized they were set to be as far apart as possible, both physically and in their transmission characteristics. This increased the quality of both. Later some packet filtering was set up to keep unnecessary traffic off the robot link. When this had been done, the system functioned much better and crashes in the last two days were rare.

Natural vs. Contrived situation Admittedly, this situation of use is contrived. However it is not clear where a natural situation for the everyday person using a robot at a distance could be found. These kind of robot control systems do not yet exist in everyday situations. Will they in the future? Perhaps. It is clear that there are more sophisticated robots in the world today then a decade ago and this trend is likely to increase.

5.2.5 The Participants

ID	Gender	Age	Profession	Research
Azure	Female	24	Ph.D. Student	Biology
Green	Male	26	Programmer	N/A
Khaki	Male	33	Ph.d. Student	HCI
Lavender	Female	29	Ph.d. Student	CSCW
Lime	Female	29	Journalist	N/A
Magenta	Female	25	Ph.d. student	HCI
Maroon	Male	26	MS Student	Robotics
Mauve	Male	42	Sr. Researcher	Robotics
Orange	Female	44	Translator	N/A
Peach	Male	30	Ph.d. Student	Film/media
Plum	Male	28	Researcher	VR-HCI
Purple	Male	30	Ph.d. Student	Robotics
Silver	Male	34	Sr. Researcher	Linguistics
Yellow	Male	56	Professor	HCI

Table 5.1: Table of Participants.

Gathering Participants The message calling for participants was sent by email early Monday morning (12:47) 6th of December 1999. The first response arrived at 8.08 and the eleventh at 13:25. This nearly filled the week's schedule of experimental runs (13 slots). The email was sent to the email lists *cid-utskick* (an email list for seminars of the KTH Center for User-centered Design),

CAS-alla (the mailing list for the Center for Autonomous Systems at KTH) and *alla@sics* the mailing list for people at the Swedish Institute of Computer Science. The first set of 11 responses from people clearly pre-selected for: people that have email, people on these email lists (mostly academic computer people) and for those that read email on a Monday morning. A few slots had purposely been left empty to recruit others from other non-academic and non-technical arenas to get broader perspectives. Of which an additional three persons were found. In the final schedule there were three slots left open. The first two were left open on purpose to allow for additional programming and setup time. The slot on Thursday night was left open because of a person who did not show up. The final schedule for the task sessions is displayed in table 5.2.

In this chapter, to avoid confusion, the persons referred to in this study are called *participants* while they are performing the task and *respondents* during the interview. The term *user* is reserved to describe that general hypothetical person that might use the system, *e.g. a general user*. A table of demographic data on the participants is given in table 5.1. Each study participants have been given an ID based on colors. This ID is meant to both preserve the participants privacy and offer an easy and memorable way to refer to the participant-respondents in the study.

Sunday, Dec. 12		
Pilot	14:00	Lime
Tuesday, Dec 14		
Session A	14:00	open
Session B	16:00	open
Session C	18:00	Yellow
Session D	20:00	Plum
Wednesday, Dec 15		
Session E	18:00	Purple
Session F	20:00	Magenta

Thursday, Dec 16		
Session O	12:00	Mauve
Session G	14:00	Khaki
Session H	16:00	Silver
Session I	18:00	Maroon
Session J	20:00	open
Friday, Dec 17		
Session K	14:00	Lavender
Session L	16:00	Peach
Session M	18:00	Orange
Session N	20:00	Green & Azure

Table 5.2: Final schedule of task sessions with respondents.

5.2.6 Research questions

There are primarily two categories of research questions being asked going into the study. These categories are functional and what could be called "sociological." The intention is that both types of questions will feed back into the design iteration process to produce better understanding of what "good design" of a

robot system is and to inform the practice of building better robot interaction systems. The functional questions center around the features of the system framework, visualization, control and interaction. The intent is to examine what functions, what does not function, as well as identify new directions for design and investigation. The sociological questions center around understanding the context of these systems and the relationships potential users have with them and how this might inform design of the system. Such sociological considerations might affect the very structure of the system itself. Depth into these sociological questions is beyond the scope of this thesis. Results here are based on observation and what participants say about their concerns and ideas for the system. The hope however is to identify and begin to understand some areas that may require further study.

5.2.7 Methodology

> *The spirit of the man, the way he feels toward things, may be difficult to measure. There is some tendency to have interviews to try and correct this. So much the better. But it's easier to have more examinations and not have to waste time with the interviews, and the result is that only those things which can be measured, actually which they think they can measure, are what count, and a lot of good things are left out.* - Richard Feynman [42].

The methodology in this study is primarily oriented to discovering the specific, the exceptional, and the contingent within the context of use. The goal is in changing and improving the way the interface may be used to perform task-oriented collaborative work with the robot. The focus is on relevance, sometimes favoring specific ideas over general principles. One of the best approaches for this is involving users in an interactive process of systems design. This study should then not be seen as a final step in the design of a system, but rather a primary step in a long quest toward the improvement of design. Study participants provide information that is examined primarily for implications that suggest explicitly, or implicitly, ideas to be designed and engineered. Some of these ideas may be subtle and trivial to implement, others may involve restructuring of the system. It is in this way that engineering is believed to be an application oriented discipline that can serve eminently practical purposes largely through systems design.

The largest influences on the methodological decisions taken in this study are the practices of *Discount Usability, Participatory Design, Scandanavian School of Design, ethnomethodological studies,* and qualitative studies in general. Discount Usability, promoted by Jakob Nielsen, makes a case of the diminishing returns as the number of users in a study increase. A large portion of the space of design flaws, improvements, *etc* are covered by working with and listening to a small number of users [80]. Participatory Design in-

volves users in a design process that includes all stakeholders, regardless of perceived initial contributions, as partners in an iterative process of creating better designs [36]. The practice known as the Scandanavian School of Design is similar to Participatory Design in that it has as its goals the interaction with stakeholders as design partners. However, it fundamentally recognizes and promotes the important side-effects of involving persons that are usually marginalized in technology-shifts (*i.e.* the promotion of worker betterment, education and workplace democracy). Ethnomethodological studies in HCI center on describing the situated experience of the user in context. Rather than describe the idealized and abstracted experience of users, workers, and environments, ethnomethodology recognizes that the practical workaday mundane experience may deviate from such idealizations and that method itself may be a subject of investigation. These deviations and exceptions might then be sought for implications on design [109, 19]. Many qualitative studies attempt to convey a picture of the use and changes in a technological experience through a descriptive detailing of that experience rather than by abstraction and generalization [114, 49]. Many of those methods are inspired by the anthropological science of ethnography. It is important however to take that work the next step and maintain relevancy by keeping the focus on implications for design.

5.2.8 Procedure

The design of the study, for the most part, is based on qualitative methods. The study collected three different forms of data, written surveys, videotape of task sessions, and audio tape of a qualitative interview.

- **Survey** The written instrument for the study was a survey in two parts.
- **Video** Video was acquired from both front and rear views of the user, the rear view included the screen display of the user-workspace.
- **Interview** A post-task interview was audio taped.

The schedule for a typical session is presented in table 5.2.8. Though this was not strict, it is representative of a typical session. Participants were first greeted, offered coffee and seated in the Grotto. While finishing coffee the users were asked to fill in the first part of the survey. In order to contextualize the task, an introduction text about the task was then read to the respondent (this text is presented in the section 5.2.1). The participants were given a hands-on explanation of how the interface controls worked. In nearly all cases the participants were given oral and gestural instructions on how to use the controls while they tested the operations. During the session, the participants were left to themselves while they executed the remote search task. The time varied considerably for each respondent. After the respondent found the last flag (most cases), or the system could not be restarted (one case), the participants

Phase	Time	Activity
Welcome	2-5	Greeting, coffee and seating.
Survey I	5	Filling in first part of survey.
Introduction	2	Task introduction text read.
Demonstration	5	Hands-on explanation of interface.
Task session	10-35	Execution of remote search task.
Survey II	5-10	Filling in second part of survey.
Interview	$\leq$30	Design interview about system.
Farewell	2-4	Offering of Cinema tickets and departure.

Table 5.3: Ordering of activities during session with average times for each activity.

completed the second part of the survey. The second part of the survey contained subjective questioned related to perceived task performance. When the survey was completed, a qualitative interview discussion took place where participants were asked what they thought of the different features of the system. The attempt was to make this as comfortable an environment as possible to promote voicing of opinions. This was limited to 30 minutes, most participants took the full time. After the interview, the tape was shut off and then many participants continued to offer comments and opinions afterward. Participants were then offered free cinema tickets and led out the front door.

Video Video was set to capture picture-in-picture: a composite view of the front and back of the user. Unfortunately because of resource contention, the video setup was taken down in the middle of the experiment for a few hours and not setup again properly to record. This resulted in the rear view being lost in the latter two-thirds of the experiments. The results and analysis here are primarily based on the taped audio post-interviews in combination with the session video and written survey. The video has been used to determine system failures occurring during a session and the time for task completion. The video will also be used for documenting the study in video format for presentation along with results.

Written Survey The written survey was divided into two parts. The first was given before the task session commenced and contained mostly statistical questions, questions about perceived abilities seen to be related to the task, and preconceptions. The second part was completed after the task session and consisted of questions related to features and perceptions of the task and the system. The questions on the survey were answered by presenting a graphical "Likert-style" scale. In this instantiation, it was seven circles in a line with 'less'

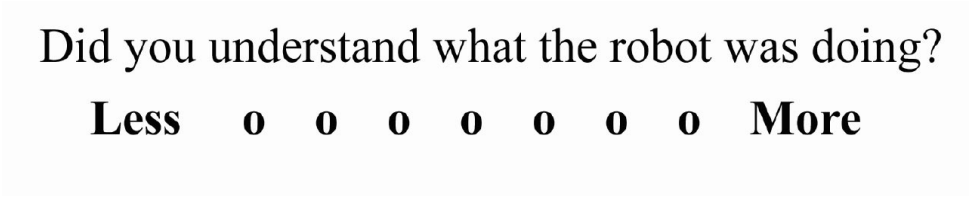

Figure 5.6: Example of survey question and Likert-style scale.

by the left side and 'more' on the rights side (figure 5.6 also see the Appendix for the survey itself).

Some overview results form the survey are displayed in two graphs. The first, figure 5.7, plots the total for each question divided up into bands for each of the written survey respondents (note that only the last nine participants filled out the survey). Each question is indexed with keywords that can be matched to the full survey in the Appendix. In this graph one can see the total and individual responses for each question. These results are sorted from high to low. One can for example see that the general perceived levels of "frustration," and "task difficulty" during the task were low and that the perceived "comfort with computers" and "task satisfaction" were high.

In the second plot, figure 5.8, the questions are plotted according to their standard deviation (deviation range: 0.6–3.0 on the scale of 1–7). Here one can see that there were consistent answers, for example, on task satisfaction and sense of what to do, but that there were relatively different opinions on how participants thought about "thinking spatially" and whether they would like to have a "robot butler."

These responses support the general impression that the participants found the system easy to understand and did not find them in periods of great frustration despite the number of systems involved and the complexity of the task.

Interviews The interviews were conducted "free-form" and were not rigidly structured (*e.g.* by a set of standard questions asked of each and every participant). The intention was to illicit comments from the participants both centered on the topics of the study and on topics the respondent wished to bring out for discussion. This represents the belief that more candid and fluid answers would be acquired with a less structured interview style. The art to this style of interview is to segue from one relevant topic to another while maintaining the respondents thread. The interviewing style in this study follows the recommendations in the informative book by Robert Weiss [118] which stresses *creating cooperation:* "an open and trusting alliance between interviewer and respondent."

All but the first four interviews where very conversational. Responses were sometimes terse, sometimes long and fluid, but always gave the impression

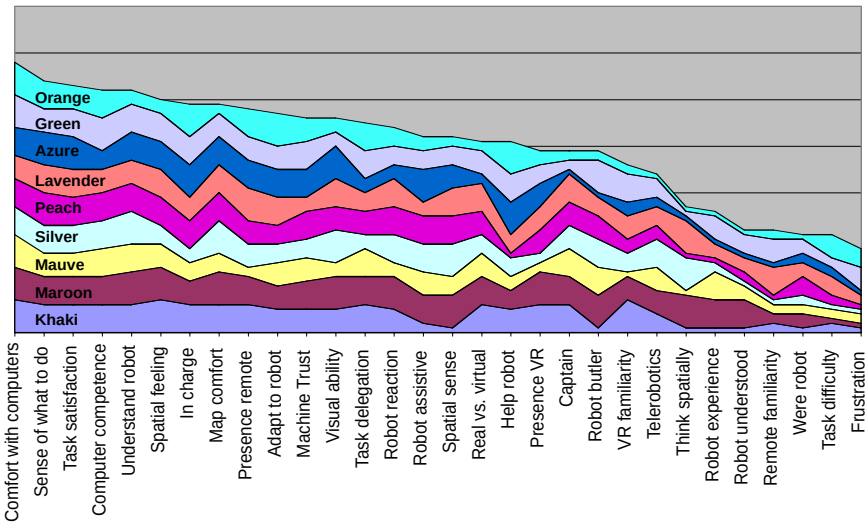

Figure 5.7: Plot of total responses, each band represents a respondent's answer to a particular question.

of sincerity. As the week progressed the skill at gently steering the respondent from topic to topic improved. A target of approximately 25%/75% of interviewer/respondent speech or better was sought in the interview, e.g. questions were short and response thrice (or more) as long. From the transcripts it would seem that this was roughly achieved. Steering the respondent from topic to topic resulted in centering the discussion around a number of salient issues which by the end became the focus of the study though initially these questions were framed strictly by the interview guide. By the end of the study these issues were refined to become the focus of the interviews. Thus, through the study itself, a number of the study questions changed and adapted as more information was obtained.

After the first four respondents it was realized that it would be better to separate out the "data" (*e.g.* age, profession, research) questions onto a written survey initially these were asked orally before the task. This would preserve the "space" of the oral interview for more fluid discussion and avoid the awkward switch from answering set questions to providing more personal responses. This strategy seemed to work and the quality of the interviews improved as a result of a more relaxed atmosphere during the oral questions. Only one person forced an end to the interview. This person agreed last minute to be part of the study and set a time before the session started of when he had to leave, thus his departure was no surprise. Often saying the interview is about at an end brought about more observations. That is to say that with the statement: "I

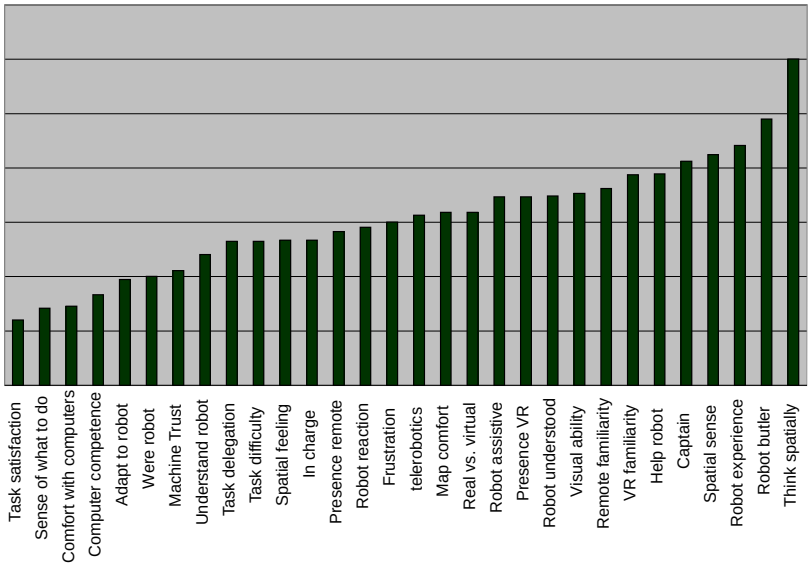

Figure 5.8: A plot of the standard deviation of the responses on the survey, the deviation range is (0.6–3.0) on a scale of (1–7). Each horizontal bars in the plot are separated by 0.5.

guess that is about it, do you have any observations or questions?" additional observations or questions produced another 5-10 minutes of dialogue.

Interview Guide To keep the interview on track and interview guide was used and employed as "crib notes. The interview guide was printed on paper, folded up and initially held in one hand. After the first five interviews the format of the interview had been memorized thus relaxing the need to look at the guide. By the last few interviews the guide was left outside the study area. The interview guide certainly "primed" answers from the respondents, but this was also the intention. The desire was to have a discussion that centered around a number of topics. The exact composition of this discussion was decided by the interests of the respondents. Respondents were never cut off, but were steered back to the topics when the discussion began to wander. This is an active process on the part of the interviewer and it is in this capacity that the interview guide is useful. The guide is presented here:

Interview Guide:

Tell respondents:

"There are no 'right' or 'wrong' answers. The main purpose of

this study is design. It is to explore the features of the system, what is good or bad, what can be improved and to better understand how people relate to such systems."

A). Visualization

Real vs. Virtual correspondence
Presence, Distance, Familiarity
Spatial feeling of remote space?
From VR world? From video world? From robot view?
Objects in virtual world?
Navigation?
Could you walk me through how you found the flags?

B). Robot Interaction

If you were to describe to someone what the robot did, what would you say?
Did the robot react to your commands?
Was the robot assistive?
Can you describe a time when it did not?
Did you understand what the robot was doing?
How well did you think the robot understood its environment?
Can you talk about how independent the robot was?
How would you describe the relationship you had with the robot?
(further: was it a partner, tool, assistant, annoyance, attentive?)
To what extent did you adapt your behavior for the robot?
Did the robot require your help?

C). Task Interaction

Was it difficult to find the boxes?
Did you have a sense of what you wanted to do?
Were you satisfied with the completion of the task?
Did you become frustrated?

5.3 Report of Study

The interviews were conducted after the participant had completed the task and the second part of the written survey. The taped interviews were transcribed (by myself) and then divided into categories for discussion. In this rest of the report the participant now takes the role of respondent and is referred to as respondent in the text to make this distinction clear. These categories are presented in the following section.

5.3.1 Discussion category overview

Fidelity and detail of the CVE: This centers on how much detail should be in the virtual model. What resolution these should be and whether or not this parameter of detail should be variable and if so based upon what factors.

Use of CVE: These comments center on the perceived use of the CVE, what is it good for, whether it is necessary, what the alternatives are.

Views within the CVE: This category centers on the usefulness of the views within the CVE. There were multiple views available, two side views, a top view, and a view from behind the robot.

Robot Controls: There were a few different ways to control the robot, these were primarily the point-to-go interface where the user clicks on the floor, and the click-arrow interface where the user employs graphical arrows to steer the robot.

Reality Portals: The RPs in the system were the flags on the box, the CRT monitor screens and the display of textures on the shelves.

Monitor metaphor: The monitor metaphor confused a number of respondents. To some it was clear, to others it required some understanding to make sense, others found it frustrating.

Camera Video: The two side screens contained camera video for half of the study. Sometimes this proved useful, other times not.

Spatiality and Navigation: Some comments were made on spatiality and navigation in the CVE.

Virtual and Real Correspondence: A number of respondents commented on the relationship between the real and the virtual both when it made sense and when it did not.

Lag and Error Handling: The issue of lag and robot breakdown emerged as a separate issue. For most respondents, when there was great lag in the interface it was not clear whether the interface had broken or if they were just waiting.

Human-Robot Relationship: This category contains information on how the respondents viewed the robot, as an assistant, as a tool, as a machine and what they said about it.

Division of Labor: When did the robot and the respondent as user establish a working relationship on the division of labor. This division appeared to be flexible and center around the robot's perceived competence.

Trust: Trust came up as an issue for some respondents. What they trusted in the system, what led them to have less trust.

Avatars: Avatars were not brought up as a concept in the study, but a number of respondents were familiar with the concept. The idea of multiple representations for the user and robot was questioned.

Speech: The idea of whether speech would be appropriate came up with a number respondents.

Box-click: Clicking on a box to go to that box was a specific new design suggestion that came up early in the study, but required more programming time than was available. However this interface feature was discussed in the interviews.

Applications: A number of applications are suggested by the respondents for the system.

The categories were formed after the interviews, they were partially set while listening to and transcribing the taped interviews. The final categorization did not change much during the categorization process itself. Some categories were combined and others were divided up into sub-categories during the collation and report writing. However no new categories were formed. This says something about the stability of the categorical decomposition of the interview tapes. Of course the categorization reflects the contents of the Study Guide used during the interview process as that was used to guide the interview questions, though the categories also changed as new questions arouse.

5.3.2 Interview Reports

In this section the respondent's voices are used as the data to discover design suggestions and general principles on which to base design of this system. When quoting respondents, an attempt was made to transcribe what they actually said. Punctuation is added as seemed appropriate during transcription. However non-verbal hesitation markers (*e.g.* um, er, ah) have for the most part not been included, but other markers (*e.g. like, well, ya know*) have been retained. When there is a long pause in speech (five seconds or longer), ellipses is added to indicate this. When there is text omitted from the speech a bracketed ellipses is used ([...]) in place of the missing text. Occasionally text is added to clarify the intended meaning that is clear in context but not when quoted. The added text is placed in brackets.

The sections that follow are presented in the order given in the category overview. Here these categories are explored in depth. In the next chapter the findings and analysis from this chapter are presented in a more succinct form.

Fidelity of the CVE

During the study the question of how much detail should be in the virtual environment was asked. How closely should the real world be modeled? This is not a matter of the precision of spatial-temporal correspondence between the two spaces, but a matter of how much of the real world is actually represented in the virtual environment.

A few major themes emerged from the discussion of this topic. One of the most important and perhaps initially surprising, is that the sparseness of the virtual environment is not a failing of the modeling, but a potential benefit when it comes to completing a task. It also emerged that this amount of modeled detail, something referred to here as *model fidelity*, should be fluid and is almost certainly dependent on the task.

The following quotation initially led to this line of enquiry with the respondents.

> [Magenta]: There are a few problems with this kind of worlds. I think it is the way we design things for it, because in the real world when we design things we have a set of constraints, like gravity, but in this world we have to design something new, we have to design gravity. [...] These are design issues, do we need that shelf? Is it there for aesthetics or is it there for some other purpose? It is like have tables, you could have designed this with the boxes hanging in the air...

When possible, a world with less fidelity was demonstrated by deleting objects from the world and asking what, if anything was necessary. Respondents answers fell into a number of sub-themes, these are *Landmarks and Navigation*, *Coherence* and *Filtering*. These sub-themes are discussed below.

Landmarks and Navigation Many participants relied on the virtual model information for navigational clues, both for themselves and for the robot. Aspects of the virtual world served as clues to location and reference. The world was not symmetrical, so there were distinct "landmarks" that served as points of orientation and location. When using the system, respondents found their way through the virtual environment by remembering certain locations.

For example the following respondent commented on the structure of the world:

> [Lavender]: For navigation, what you need is landmarks, so it wouldn't help if you had shelves all around, or walls. The shelf that sticks out, this a landmark, and the globe.

The fact that there was distinct locations in the space helped the user orient herself and navigate the space. This suggest that if the real space were symmetrical, one should make an effort to add features to the virtual model

to make a distinction. In fact, the "globe" mentioned above is such a feature. It was bright texture on large projection screen that exists in the lab, it was turned on in the virtual model, but not in the physical lab.

Maroon asks for a simplified world that does not contain the textures (Reality Portals). Or at least not the Reality portals that are not related to the current task. The textures on the flag boxes were also Reality Portals and it is clear that one must see those if one is going to use the model for the task.

> [Maroon]: I think with tables, bookshelves, you need them for orientation, but the texture is not that important. I think it actually helps when you have the basic structure plus the information you need, like the flags.
>
> [Question]: You could put a lot more in the world, what is necessary?
> [Peach]: Now you are talking about landmarks, and such stuff. Yeah I think that the shelf here is, of course, a landmark, but I don't look at the flat surfaces, I did not look at my own picture there. I wasn't looking at the surfaces, just the objects. But then the fact there are no things on the table here. The task was to find these rather large objects. But if I wanted to find a book on the shelf, maybe I would need to see the details of the objects on this shelf. It is all about scalability I think.

Here Silver claims that what you want is "a lot" of information in the virtual space. However he also only talks about the boxes.

> [Silver]: For this task you need a lot because you want to find the boxes and you need them to be able to navigate around to them.

This suggests that certainly you need the boxes, but do you need anything to support them? When possible the respondents were asked what they thought about a world with only boxes visible. Would that work for navigation and for the task? A number of times such a world was demonstrated by deleting all the objects except the boxes.

> [Question]: What if boxes were just floating in the air [demonstration of floating box space]. Say you had nothing but these six boxes floating in space?
>
> [Mauve]: Yeah, you need some kind of boundary, if you just have this blue background, you can look around in all directions it will take you some time. Especially if you didn't know where you had been before. You could rotate around many times and not know where you were before. [You need] some kind of reference point.

This relates back to the notion of needing to have reference points. Especially if an understanding of the structure of the real space has already been established.

> [Mauve]: It depends on the task. if you were doing this kind of task, you would just want to see the boxes and perhaps you would like to see some other things that have to do with navigation. "How do I have to go around things?" etc, had you just showed the important information it would be more simple to do this.

Mauve raises the point that the fidelity of the world is dependent on the task and that the representation of the world is also important to the establishment of the user's understanding of the robot's working environment. This is an important finding, and makes strong suggestions for the design of this system. This issues is raised again in the discussion section.

Coherence For some respondents the lack of structure (walls, floors, tables) in the environment caused confusion. This confusion impeded their perceived ability to use the space for the task.

> [Peach]: Then it is a matter of spatial immersion, you can have these six boxes floating in the air, but then you would not have the feeling, or experience of the room. [...] you need a ground to stand on. Flying fucks up your gravity and things, [a] floor is important.

Peach implies that certain elements from the physical world should be brought into the real world to avoid confusion. Most virtual environment systems rely on the ability of a user to take some of their "dirt world" knowledge and use it in the virtual space. For some this brings up a notion of "natural" interaction:

> [Lavender]: Then I guess you wonder "how on earth are the boxes floating?" because in the real world the boxes aren't floating, so maybe someone would like to see how they are attached and you don't have to bother if the table is there, because that is natural, more natural.

Bringing in elements such as structure (e.g. walls) and physics (e.g. gravity) provide a basis for assimilation into the virtual environment. In the Flag Task, the respondents did not have much opportunity to practice. It is not unimaginable that other successful metaphors for navigation exist in virtual environments that do not rely on, for example, walking through cartesian space. However it is almost certain that they would require some training. In this context however, the use of real world notions of spatial navigation served to facilitate quick use.

Here Orange mentions that she feels more comfortable in a space with things that are "familiar," in this case about the video screen.

> [Orange]: There is a lot more information. For me that was good, and the information was much more familiar. I am much more familiar with being in a room full of things, than moving around a graph.

Such a notion of familiarity is related to the notion of personal comfort and is certainly a subjective quality. This has the implication that the amount of fidelity in the room should be user-configurable. That is to say, fidelity should be adjustable to the comfort level of a particular user.

Magenta suggests that if one were already familiar with a space, then less information might be needed. However that there is also some "loss" when objects are deleted.

> [Magenta]: Of course I would feel like loosing something. [...] If I would have had a map, and if I could have been in this room, to be able to see the room in my head.

Magenta says that having a map and the room "in her head" would help.

Filtering The notions of landmarks and coherence were in potential conflict with a desire to filter out information by the users. While the above sections discuss the need for a certain amount of detail in the model this section explores the other situation when a user wants to filter out information. There are times, most probably when a user is highly focused on a task, that it would be helpful for certain elements of the model to be removed. A case for such a feature in the system is well put by Orange:

> [Question]: What do you need to see?
> [Orange]: Depends on what your goal is, if you have a limited task, then you can eliminate the extraneous information because you can focus on what you want to do. [...] I could imagine if this were in a real context. I could imagine having a complicated room entirely programmed into the computer, and then having different filters. Right now, what I am interested in is books and have the book filter on. Another example is "all the lamps" and everything else disappears.

Thus filtering may be task based. In such a case one might have a rather complete model available and only visualize parts of it at a time depending on your current needs. This sort of filtering may be something we already perform:

> [Mauve]: As a human you are used to take away things that are not necessary to solve for a task, and I think this could assist in some sense, you get rid of some things that disturb.

In fact there is research in the vision and computer vision community that builds on precisely this feature. The idea that we have a form of selective perception available to us and artificial systems have been built to take advantage of this [40].

In addition to the fidelity being dependent on the task at hand, it may also be made to be dependent on robot competence. There may be obstacles in the real world that do not require visualization because, for instance, the robot the robot can competently navigate the space. [Orange]: It is probably not a good idea to get rid of the bookcase as that could be an obstacle." If however the robot failed to navigate around a particular obstacle, it may wish to display the obstacle to the user. In this way some of the filtering could happen automatically.

> [Silver]: So you want to say, "scratch everything but the boxes" take all that out and leave the boxes. "but if you see this mail in my mailbox, I want that anyway." I am thinking of this like a home service.

In fact, some respondents implied that this filtering should and might already be happening automatically (it was not).

> [Orange]: I would assume that in this overhead view that you have eliminated extraneous information, so in point of fact, the boxes there are the boxes that have flags on them.

The system for filtering and providing levels of model fidelity has the potential to greatly enhance the system, increasing the flexibility for different task domains.

User focusing, filtering As mentioned, task-based filtering is something humans may perform naturally. A number of users reported rather selective vision when asked about the details of other elements of the virtual environment.

> [Question]: What about the other textures in the room?
> [Lavender]: Well um.. I don't think you are conscious of them. [...] I didn't bother about those [virtual monitors], because I figured they didn't have any function for me. I didn't think they had any information for me.

> [Question]: Did you notice those [RPs on the wall]?
> [Green]: No, it was only boxes for me.
> [Azure]: ... because that is what you are looking for.

> [Question]: You were distracted by CRT?
> [Mauve]: No not really because they were a different size, I think I went from size and number.

> [Question]: Did they look like CRTs?
> [Mauve]: Yes.. [chuckles] ... I saw this image before and it was quite natural to think these were computer monitors.

Different users employed different parameters to perform the search task. Many agreed that they simply did not "see" the elements of the model. The boxes popped out, partly aided by the bright texture color.

> [Yellow]: I was quite concentrated on the task of exploring the flags, of course the real space is much more decorated with cables and all the computers around. I did not look so much on that, I was much more concentrated on where the maps [flags] were I now realize that there are much more details. I probably mostly used the virtual presentation in my navigation.

Yellow realized the image of the real space in the video contained was more "decorated," or rich with detail, and that surprised him. In fact given the choice between the richness of the real video and the impoverished virtual environment, he chose to use the virtual space.

Use of CVE

One subject of query in the study was to question the role of the virtual environment. Is it necessary? What affordances does it offer? What are the qualitative differences of having it?

During the discussion, the screen where the virtual environment was displayed could be turned off, or the world could be deleted entirely. The respondent could then use the system without the CVE to control the robot. Although this was not done in the context of the task, it did offer a chance for comparison. Most of the responses centered around the need for the top-view, or map view. A more in-depth discussion of the role of the different views is taken up in the **Views** section. However, it can be said in summary that the top-view turned out to be the crucial view of the CVE. The following respondents make reference to performing navigation and localization using the virtual environment.

> [Question]: When do you use the virtual world instead of video?
> [Khaki]: This [CVE] is where I am doing the navigation, then I peak over here to look at the next flag.
>
> [Green]: Yeah, but that [CVE] very much helps to know where you are.
>
> [Question]: Is it helpful to have the virtual world then?
> [Mauve]: Yeah, for me, I use it all the time, if I could [also] look here [at the camera view], but I think the VR world was ah.. this overview image gave me a lot.

> [Question]: What if you did not have the virtual environment?
> [Lavender]: But could I point in the video at the floor? Because the virtual world is ah, well, it was quite good to only be able to specify a place where it [the robot] goes.

Here Lavender brings back the notion of using landmarks to find herself in space.

> [Lavender]: ... in the few first seconds it was quite hard to know where things were in this space. But I guess the shelves and the flags made it quite obvious.

Some modeled features in the virtual environment were not in the real physical environment. These were a graphical display of the boundaries of robot travel, an arrow indicating robot direction, a semi-visible cone that indicated the camera viewing angle. Here Lime makes reference to the usefulness of seeing the robot's boundaries: "Without the virtual world you lose sense of direct relationship and loose the boundaries."

CVE compared to Video A number of respondents compared the virtual environment to the video. The video screen on the right displayed video of the room from a particular fixed location. From this location roughly 85% of the task space was visible. One of the questions asked of the respondents was whether it would be possible to use video alone.

> [Magenta]: But it feels strange using this here [respondent experiments by moving the robot without CVE]. I think it is difficult because you don't get the feeling of where the robot is. It is black [the robot], but it is hard to see it [in the video], it blends with the background.
>
> [Plum]: I could see them [the flags] in the virtual environment. If I had to go only from the 2D video maybe it would have been harder.
>
> [Mauve]: But I see now that guiding the robot through 3D is very difficult with the video. You might have to have several views if you just see what the robot sees, you will not have enough info to guide an arm.

One approach to making the video more spatial is the use of stereo video, thus offering the user of the system a form of depth perception. Maroon did not see that adding a stereo camera would help. A stereo system would certainly introduce new problems as well.

> [Question]: Does the idea of [having] a 3D world give you a sense of the space?
> [Maroon]: Yes it does work, it gives you more of a sense of the space

> than the camera would, because it [the camera] is too narrow.
> [...]
> [Question]: Even when it is not [in] stereo?
> [Maroon]: I don't think stereo would help, we are used to analyzing pictures like this and seeing 3D information.

Not all users responded positively toward the virtual environment. One respondent did not feel comfortable using the "graph," and felt that the video was more familiar.

> [Orange]: Once I have the camera views available, this virtual environment view became useless to me, well I don't know, I no longer used it at all, it may have been helpful if I had thought about it some more or had more time to play around.

Another respondent suggested a system that would combine video and the virtual model into a form of Augmented Reality. In such a system information from a virtual model can be used to augment the video.

> [Peach]: Why do I need this simulated world? Because it can guide you, and it could place labels on all objects, in this environment [video] it can be hard to see what is this is from a distance, "oh what kind of machine is this?".

Aesthetics of the CVE More than one respondent commented on the aesthetics of the VR space. A number of comparisons were drawn with video games. As many video games use the same graphics library (OpenGL) as this system, it is not surprising that some would notice a similarity.

> [Orange]: I don't play 3d video games.. and I don't particularly like 3d video games when I tried to play them.
>
> [Purple]: This is a kinda video game.
>
> [Magenta]: It is a bit frustrated. I have been in a project before doing a world in DIVE. Even if it is top state of the art, it is slow anyway. Everything is so square and difficult to make round shapes.

Feelings regarding video games can be strong one way or the other. One could guess that for users with a predisposition toward video games, that, at least initially, some of those feelings would be transferred to the virtual environment and vice versa. Thus in presenting such a system to users, and perhaps in the actual design, care should be taken to consider this effect with respect to the intended audience.

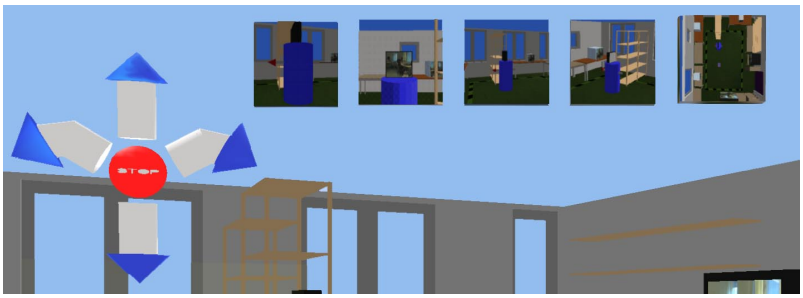

Figure 5.9: Different controls and views in the interface. The set of arrows (click-arrow) on the left controls the fine motion of the robot. The picture-buttons in the middle switch the users point of view.

Views within the CVE

This section centers on the usefulness of the views within the CVE. There were multiple views available, two side views, a top view, and a view from behind the robot. By far the most used views were the top (plan) view and the behind-robot view. The side views, for this task did not seem to offer much to the users. The side views are the views used in most work in CVEs, it is an interesting result if the need is different in the CVE with the robot.

Top view The top view affords a user to look down on the scene and see the robot in the context of the space. This was a plan view quite similar to a map. However it differed from a map in that it was not an straight projection, elements of 3D structure and side surfaces as well as top surfaces are visible depending on the perspective. The top view helped with orientation:

> [Lavender]: I guess when you were in this view[top], you could see which was left or right in the room and get your location and the robot location.
>
> [Khaki]: This overview is very effective for giving me a spatial position. [...] In the real world I have a very wide view optically and can look around quickly which is very hard in virtual space. [...] I need to look around more, to get a feel for the environment and I think I get that in the top view and there is a point to having a map when moving in a large space.

Thus somehow the top view is able to give an overview of the space to the user. Seeing this view affords a sense of the space that the other views lack. As stated above, there may be a greater need for this in the virtual world as the angle of view is limited as compared to what it is in the real world.

Some respondents indicated that perhaps they could perform the task with only the top view.

> [Question]: How would you compare the different views?
> [Peach]: [T]his was the most useful one, this overview. Partly because of the speed or the slowness of the system it makes it like, maybe you don't need other views? the side views? No I don't think so, if you know the room well enough, this overview is enough, that is maybe just my preferences.

If the user does not need the other views this may put the role of the CVE into question. So this poses the question if a map will do?

> [Question]: It is not exactly a map?
> [Peach]: No it is something more.
> [Question]: Does the perspective destroy the sense of a map?
> [Peach]: No I think it helps, because you see this is a shelf for instance. If this would just be a block, it would impede the navigation. That this is 3D is a good thing.

When the virtual world is viewed from the top view most of the structure is parallel with the line of sight. However because of the perspective some of this structure is visible. Mauve makes this point: "If you start to vary the height [of the boxes] than it might be more difficult with just the top view." For example, a number of the flags were identifiable in the top view because their front faces were visible from the top, or enough color of the flag was visible to make a guess.

Many respondents use the top view in combination with the behind-robot view. Thus they would get position and global spatial information from the overview and then switch to the behind-robot view to look at the local information on the flag boxes.

> [Maroon]: For global navigation, the top view is the easiest one. when it goes to the final tracking, it is probably the view behind the robot is the best one.
>
> [Azure]: It is a great help with the virtual picture, the top view, where you can see the robot from above and sort of guide in the right direction, and then you snap to the picture [behind robot] to see where the box is.

Behind-robot view This view afforded examination of the robot's camera video as well as a view of the world from the robot's perspective. The information in the video was necessary to perform the task, *i.e.* to see which flag was the next to visit. However when the video screens were on there was an alternative available to using the behind-robot view, the left display screen.

> [Purple]: This one is useful for navigating [top view], this one I needed for small flags [behind robot], this one I used to see more of the boxes [side view].
>
> [Khaki]: since I needed to see the flags, to me I definitely wanted to have the robot centric view.
>
> [Orange]: [I] definitely used that one, the robot's view, the most.
>
> [Plum]: I did not think the rear view or the door view [side view] were very useful.

Plum did not find the behind-robot view useful and instead used the top view and the video screen. Because of the stated importance of this view by many respondents, one can then imagine a tighter coupling of the behind-robot and top views. One such coupling was tried during the study, this was by adding a second visualization of the robot's video on the visor of the avatar. Thus up with the buttons the same video found in the monitor metaphor was displayed attached to the user's point of view. However for at least one respondent this provided an additional confusion. That of thinking that all the buttons were video if the one was. This was not a hard mistake to make, as the icons for view switching were snapshots taken from the views. This video icon was larger than the others, but if it were made more distinctive, e.g. looking like a television set, the distinction between still and video might be more clear. As will be discussed in the section on the *monitor metaphor* the quality of the video in the CVE needed improvement.

Side views The side views and the robot-centering button were not used much by respondents. These would be the traditional views offered in most CVE applications. The robot-centering feature was offered in case the user navigated their avatar away from the workspace and lost focus on the robot. It was never the case during the study that this feature was necessary. Had the task been more exploratory in nature and covered a greater extent of the virtual space, this might not have been the case.

> [Yellow]: I used only two views: from the top and from the top (back) of the robot.
>
> [Question]: Are two side views enough?
> [Purple]: Probably if you had three you could cover the whole room for the purposes of spotting the box.

It may also be that other tasks would require the side views, *e.g.* if the target items of a search were positioned underneath objects and thus would not be visible from an unaltered top view. Yellow puts it best. It may well be that the side views were not necessary, at least for this task. Thus the choice of views should perhaps be configurable and dependent on the task.

> [Yellow]: It worked so well, I really didn't use more than two views. Maybe those two views are enough. Probably this one about centering the robot would be useful if it gets out of focus if you come into that situation. But otherwise I am not sure about the other two. Of course with Occam's razor-knife you should perhaps cut them out.

Camera control When switching between views there was a zooming feature that offered a "smooth" transition from one view to another. This was implemented because of previous research with zooming interfaces that showed that users were able to maintain spatial sense when they were zoomed to a new location, instead of snapped [13]. However this feature needs to be implemented with care. Peach offers some wisdom from film studies that would be useful when redesigning the camera zooming feature.

> [Peach]: The camera movement was smooth, but the graphics were bad. Did that help? Yes, if you just flipped from one to the other. You could also think about automatic camera control. When you move from one end of the space to the other you could have automatic camera control to follow the robot. You know the robot is exiting to the left, and coming in from the right, and that would have the same effect as the cinema, using the 180 degree rule. That is the fact that if you exit one frame left, you must also enter from the right. If you enter from the other direction, you suspect "oh he came back". It's called the 180 degree rule because the camera is not to move over the line between two characters. There were many blockbusters before 1917 that were really messed up spatially.

The camera movement was disturbing to some. It was not incredibly sophisticated either. When a user selected a new view, the idea was to smoothly translate and rotate to the new position. However it did not take into account objects that were in the path between camera viewpoints. Because of this a user might travel through part of a bookcase from one viewpoint to the other. This was initially a disturbing effect. That condition was minimized by consideration when placing the fixed viewpoints in the CVE but was not eliminated as, for example the behind-robot viewpoint was dynamic and depended on the robot's position. If this is to be useful, more care can be put into the sub-system that handles this transformation and take into account virtual obstacles along the way. This is of course not trivial. Also further investigations into automatic camera control can be investigated. Many 3D video games, are able, through careful design, perform sophisticated camera control that create a seemless and enhanced experience. Some of these techniques might prove useful for task-based environment visualization as well.

Robot Controls

As described in the previous chapter there were different methods to control the robot, these were primarily the point-to-go interface where the user clicks on the floor, and the click-arrow interface where the user employs graphical arrows to steer the robot (figure 5.9). Respondents quickly discovered the benefits and uses of the different control strategies. As the following quotes demonstrate, the floor interface was used for traversing the room and the arrows were generally used for smaller local adjustments.

> [Lavender]: I guess those [arrows] are for fine adjustment. When you need to specify the view, kind of, but when you want to navigate, it is better to just click on the floor.
>
> [Yellow]: When it comes to fine tuning, left/right [arrows] is much more useful. Come close to the object, bring it into focus. When it comes to moving large distances of course this teleporting [point-to-go] makes it much easier.
>
> [Peach]: I use this distance clicking a lot. Ok I want to see this object and just click here and just let the robot go there.

Command mode comparison The floor interface was particularly useful from the top (plan) view of the room. In this view the extent of the floor was generally visible and the boundaries for robot exploration were made clear. Also the spatial arrangement of the flags was most obvious in this view. For those that had a good sense for maps this was the most useful way to navigate the larger spaces.

> [Yellow]: This top view helped to get a view of the room, and quite useful was this to point at the floor - it was a very nice feature. Compared to steering the robot all the way, being able to teleport yourself in that way, it was very helpful.

Yet Green enjoyed having remote control of the robot. As Silver eludes to, this may just be initial fascination, and eventually what you want is higher-level commands:

> [Green]: I liked this one very much [the arrows], just move around where you want.
>
> [Silver]: Driving it would be fun, but [...] I would like to give a command like "ok circle around the room and look for flags." [...] I don't want to actually manipulate the robot. I would like to tell it, give it macro commands. I noticed you had questions about trusting and delegating, I would delegate that sort of stuff.

One of the problems with the arrows interface was that it could be tedious to use. The way it was designed was that a click would generate an increment of movement. This was chosen because it made the control communication easier between the robot and the virtual robot agent. The following comments identify this problem:

> [Magenta]: So then I started to use the clicking on the floor, and there it was one command, go to that place. Using the arrows, bump bump, is very slow.
>
> [Khaki]: Also this is something I really would have preferred to have continual control over, this mouse moves in small increments which may be too small or large. Also when you just want to move in a given direction for a certain while it would be more convenient to just press on th[at] [button].

Because of this comment a new method for continuous movement was implemented. Between sessions this second way of commanding the robot through the arrows was implemented. If you clicked on the body of the arrow you moved an increment, but if you clicked on the extremity of the arrow you moved continuously until you pushed the center stop button. Another possibility would have been to move the robot while the button was pressed. The biggest problem with this is that because of the inertia of the robot, the response of the robot stopping will lag behind the release of the mouse button by a few seconds. The feeling was that providing a click-and-hold-to-move interface would provide false assumptions on the control the user had: that if they released the button the robot would stop immediately.

Continuous Robot movement After the new control was implemented a number of issues of controlling a large mass that has inertia and does not stop immediately manifested themselves in different ways.

> [Lavender]: I think that they move too fast maybe. I felt always in a hurry to stop it. ... It just felt like it didn't stop sometimes when you wanted it to. [...] If you could adjust it so you could decide how fast it goes?
>
> [Maroon]: Speed is ok, but the turning is a problem, probably it should not go that fast when you go the continuous speed. Because there is too much inertia in it, and you are not able, at least on the small video, to detect things [as a user]. When you have a "go-point" behavior it is ok to go fast, but when I am trying to do small movements I want slower.

This is also an argument for having variable speed. However it was believed that there was probably a good compromise speed for the continuous movement. After a few experiments with participants during the study a good speed

was found, however it was not constant. The robot motors have two parameters that determine the speed: acceleration and velocity. The acceleration and the velocity can be set separately for both the rotation and the translation speeds. The speed used was just above the threshold of being too slow to be frustrating, but just below the threshold of being too fast for most people. This speed was set to approximately $20CM/sec$ which is about half the normal translation speed of the robot. To accommodate the desire to cover small distances slowly and larger distances more quickly (both angular and translational), the acceleration was set to be relatively low. This gave the behavior that the robot would start slow and pick up velocity to its final set speed. This behavior turned out to seem quite natural as the robot would cover small distances slowly and then speed up the longer it was in continuous mode. The rotation velocity was set to be slower than the translational velocity. The continuous rotation was generally used for panning around the room looking for flags. While in this continuous mode the collision avoidance system was still operative so the operator does not have to use the stop control to avoid obstacles.

Competition for visual focus One of the comments implied that the graphical arrows may have been misconceived. The reason for this is that both the task and the arrows require visual attention:

> [Lavender]: One of the problems is that you need vision to use these and you need vision to do the task. So then you go like this: "am I on the stop now oop oops ah" you are checking here, "am I hitting the target?" and here you have to stop the spinning.

The suggestion was to consider a joystick. This is a good point, as the joystick would not have this contention for visual focus. However the joystick may not offer methods for specifying the deictic references, such as pointing on the floor, as easily as a mouse or glove. It would also not be an improvement to clutter the user control surface with various devices, joysticks, mice, etc. However if one had a hand gesture system to track pointing at the screen, a joystick might work as the one device for fine control and if communication delays could be minimized, this might function well as the lowest level interface.

A new control interface that was implemented as result of this comment which was to click on the video within the virtual environment. Clicking on the middle of the image made the robot go forward, clicking on the sides made the robot move to the sides. In this way it was as if the user clicked on the real objects to go to them. Few respondents had the chance to use this however.

Reality Portals

The Reality Portals, as described in the previous chapter, insert video views into the virtual world (figure 5.10). The RPs in the study system were the flags on the box, the monitor screens and the display of textures on the shelves.

Figure 5.10: A view of the Virtual Environment model used in the study with a few of the Reality Portals in view.

Attempting to evaluate the functionality of RPs in this context is difficult. The RP behavior was such that they did not become visible until the robot had 'seen' them. One example is the textures on the flag boxes which did not become visible until the robots view cone had passed over them. These were of course seen by the respondents, but were not not noticed as "video textures," but as part of the environment. This may in fact say something rather outstanding about their nature, that they were seen as belonging in the scene. In the discussions it was in fact difficult to elicit comments on the reality portals in particular. Also the flag textures and the shelves were not seen as in the same category of display device. One reason for this was that their granularity was different than the other RPs in the environment.

> [Question]: What is the difference between flag textures and bookshelves?
> [Lavender]: It is another texture, a different granularity, it is blurred.
>
> [Peach]: The colors, yes, these flags are very clear and colorful, so you know. Also these colors are more prominent than other colors in the room, they pop out, they are stronger.

The comments that were made indicated that the flag RPs were vital, but the other textures, bookshelves, *etc.* were seen as clutter and distracting.

> [Mauve]: I think pictures help if you can get them, it gets more real in some sense.

> [Silver]: If I really was using the robot for something real, I would, when I get to see this thing, [..] look up information and share this information with the robot. And this lighting up [RP turns on] means something. while I am doing some other task, the robot could leave information, so I could take advantage of that.

Some respondents then saw the RPs as information resources, as well as contributing to the sense of pictorial realism. However some of the textures were strangely placed. One in particular contained a good deal of 3D information, and because a user could go behind it (it was on a shelf in the middle of the room) it appeared to be floating in space from some angles.

> [Azure]: Yeah I think it is ok to have it like that and it will be particularly strange when you move around, when you get closer and it changes.
>
> [Lavender]: one thing that is strange is that picture on that shelf, it seems like a poster, so I don't understand why the picture is there. Oh.. it doesn't resemble the junk from this angle. I don't think you need it. Now it is harder to see it is a shelf.

Orange, one of the less technical respondents, clearly envisioned one of the intended applications which offers some suggestions about possible domestic use.

> [Orange]: I guess that would speed things up if I don't have to send the robot there every time, it is like having surveillance cameras in different parts of the room or rooms of the house, and then you can survey the space, and then decide where you want to go.

The quote below demonstrates that when RPs are working they work well. Until the robot passed by the RP locations they were not visible. In this way, participants experienced both having RPs and not having them. Here Purple does not make the distinction that the boxes are RPs as well. He saw flag textures as something necessary but the other textures as unnecessary.

> [Purple]: I was actually happy not to have textures, because then you see the box immediately.

It is hard to conclude implications on design, except to say that clearly the desire for RPs in the CVE are also relative to the task. While the flags were not seen as textures, other elements not vital to the task were seen as unnecessary.

Monitor metaphor

The monitor metaphor confused a number of users. To some it was clear, to others it required some understanding to make sense, still others found it

frustrating. This confusion centered around two issues, the quality of the video in the virtual world and the placement of the video in the monitor on top of the robot. The video in the CVE was low quality and is clearly something that needs to be improved. The monitor metaphor did not quite appear as "natural" to some as was presupposed. Although many understood it, some had some difficulties getting "passed it."

Video Quality The video quality was noticeably lacking, it was small, it lost color resolution when it was brought into the CVE and noise was introduced by the radio AV link used by the robot. One respondent commented that it looked like they were looking through a "scrim[1]"

> [Lavender]: bad quality, but I guess it is enough to recognize things and where you are.
>
> [Magenta]: I think the picture is too small to really see the flags. When you get used to it perhaps it would be different.

Khaki saw the video as a way to see the real world inside the virtual environment as the robot panned around the room:

> [Khaki]: It was more like the video view was the "high resolution" version of the low-res graphics.

Video placement Some of the respondents comments centered around the strangeness of monitor placement:

> [Magenta]: It feels strange that the robot has eyes in the back, that the robot has a screen on it showing backwards. That feels strange. As it feels like eyes.
>
> [Question]: When you see video from robot's perspective [in the virtual environment], does this makes sense to you?
> [Orange]: Well, no, it doesn't really make sense. What I find confusing, is that when I look at it on the screen to my left [robot camera], then I understand that I am seeing what the robot is seeing and therefore what the robot is facing. When I see it on the robot itself [in VR], for some reason, I don't understand why, it is not logical at all, but when I see it on the robot itself, I feel like the view that I am seeing is the view that is behind me in the experimental world [...] I cant get passed that there is something blocking my way.

[1] A semi-transparent cloth often used in theater to create a special effect atmosphere.

The above was a new realization, the belief had been that the monitor metaphor was somehow natural and clear. However this was one of the reasons for including non-specialist users in the study. Most previous viewers and users in demonstrations had been "male engineers." Bringing in respondents from the outside, brings different perspectives.

It is clear that the quality of video needs to be improved in the CVE. It is also clear that more thought should be given to placement of live video. There are of course, alternatives or choices offered to for this video placement. Another placement would be to situate the video at the end of robot's viewing cone, so that the video appears beyond the robot and appears more like a special lens on graphical world supporting Khaki's statement above more concretely.

Camera Video

The two side screens in the grotto contained camera video (figure 5.5). The screen on the right contained a view from a fixed camera position in the room. The screen on the left contained the video from on-board the robot at full screen size. Neither of these video images needed much explanation and their presentation was clear by moving the robot. If the robot did not move, the scene was relatively static. Only one user reported being confused initially by the video views. These views were displayed during the explanation of the task and interface and then turned off for the first half of the study. As soon as the Tanzanian flag was found, marking a soft half-way point, these projection screens were turned on. Conflicting reports of the usefulness these video displays were received. One question in the study was whether they were necessary or if the CVE itself would suffice.

The video quality from the camera, although mostly clear, was not fantastic. The color contrast between the objects in the lab was low. This was not a technical problem as much as it was a problem with the environment. Here Lavender makes a comment related to that limitation, and Mauve comments on the limited field of view:

> [Lavender]: From these views, I don't think I would be able to help it [the robot] ... I mean it is too low sharpness so I cant see whether it is bumping into it or not. Also that is not 3D it is 2D, so it is difficult to see in depth. Perhaps if you start with better quality, it could be easier to see where it is in relation to the objects.
>
> [Mauve]: Some of the questions [from the survey] are about the camera views of the robot and I see now they are hard to understand, no not to understand but let's say spatial information is difficult because they tend to have a very limited field of view. You feel a bit blind in some sense, you see just this little window and you wonder what is to the side and when I see it here it is quite obvious that is is very difficult for a human to have just this small look. I

> think when you are a kid you are trained with this TV games, video games, you are fast in turning around like this, but I am older, I find it annoying and don't see what is to the left and to the right.
>
> [Question] Did anything change with the camera view [on]? [Yellow]: Not much, you feel a bit more comfortable and it is nice to see the real robot move together with the virtual robot. I have the ability in the VR world enough, it was not something I needed to use with navigation.

Mauve touches on a limitation of video that motivated the idea of a Reality Portal. That is one of the spatial limitations, that you can't see what is outside the current video frame. Panning around the room to get this sense takes time as Maroon points out:

> [Maroon]: When I use the camera on the robot, I would have to do a kind of scanning first, for myself to get to know. I probably get lost once in a while, you know, when I do this 2D scanning, that for sure takes much more time.

Orange found the robot's view (from the on-board camera) confusing. This is related to the confusion she had with the monitor metaphor. After moving the robot for a time, the meaning became clear.

> [Orange]: When I figured out that the camera view was behind the robot, then that was good in order to just, first to know where to go. [The] "camera view", it was helpful to have another angle in the room, because you could see if there was another object you were interested in.

The task could be performed without the side screens. Because of the degradation of the video inside the virtual world, the robot's view as displayed on the projection screen, was preferred to the one in the CVE. The initial answer on whether a user could use just the CVE is that because of the poor quality of the video in the CVE, it would not be preferable yet. The video in the CVE would have to improve, but then it might be possible. The first three flags were found by all users without much difficulty before the video screens were turned on.

Spatiality and Navigation

Some comments were made on spatiality and navigation. For most of the participants, navigating around the environment presented few problems. However most of the participants were also male engineers, who stereotypically have few problems with maps and orientation. When Peach says "I don't have problems in know what's right from left, etc," that statement common to how many responded on the survey to the spatiality question. However for others orientation and moving around, especially involving rotation, was more problematic:

> [Azure]: That was difficult, not really knowing if I had turned, how much I had turned, do I have my back to what I had my front to a minute ago? [...] I mean I usually get lost, as I don't know left from right. [...] I know when I was trying to learn to drive the car, and they wanted me to go backwards, and to turn when I was going backwards, I would always turn the wrong way.
>
> [Orange]: I had no idea where I was. So I had to figure out, "Ok, why don't I just do this: if I push the forward, and this thing gets further away from then I know the whole thing is behind me." So it was like I had to figure it out, I didn't get a sense for what this room looked like at all.

It is not clear how to build a CVE that makes these concepts easier. For many the top-view worked fine for spatial orientation. For the others, where maps are not as effective, one might try a strategy of adding landmarks. That is to say, if the environment lacks landmarks, it would be possible to add virtual beacons that help orient the user in the world. These beacons could be ambient objects that fade to the background when not attended to. An example of such a feature might be coloring the walls differently depending on their orientation in the room. Thus one might be able to passively absorb that they are facing "virtual north" because they are looking at the blue wall.

Virtual and Real Correspondence

A number of respondents commented on the relationship and correspondence between the real and the virtual environments. At times this correspondence was tight and made sense, and a few other times inconsistencies or predispositions detracted from the correlation.

For most of the respondents the correspondence between the virtual robot agent and the physical robot was clear. For most, when the camera view came on seeing the physical robot did not startle the respondents and their conception of the robot did not change.

> [Khaki]: I don't think I made a distinction in my mind between which was the physical robot and which was the virtual robot.
> [Question]: Did the camera view help you understand there was a real robot?
> [Khaki]: No I thought that was pretty obvious, but then again, I have seen this robot. I had no problems with that [the correlation].
>
> [Yellow]: Of course the real robot moves also a bit smoother than the virtual robot did, but otherwise I saw the correspondence clearly.
>
> [Lavender]: Of course I could see it was corresponding to the real world, there were the same number of monitors.

For one respondent, switching on the video was surprising mainly do to the lack of correspondence between the virtual robot position and the real robot. This lack of correspondence was caused by sloppy initialization of the robot position. A rush to begin the study session made it so the robot was not positioned carefully enough on its start marker.

> [Purple]: When you switched it on [the camera video], then I was a bit confused, the robot standing in front of Chile, and I still had Hawaii so I got a bit confused.
>
> [Question]: Did anything change when the video came on?
> [Magenta]: Not really.
> [Question]: Did it confirm anything?
> [Magenta]: yeah, yeah, hmm... I think, I don't know, I didn't know that you, I was not thinking you had a robot running around - so it was actually a robot?
>
> [Purple]: Was like computer game to me. That was a bit funny when you switched that [video] on, it was like "oh there is a real robot moving". So.. but to complete.. to do the task or so, I did not care if there was game or if there was a robot.

For Yellow, and probably many others, it was when the software system broke down ("crashed") that the impression of presence vanished.

> [Yellow]: When that [robot connection] broke I really did sort of get out of the space. I did not get into by just, by looking at the live video as I could not navigate.

Plum did not feel that it mattered where the actual physical environment was, it could have been very far away, or in the next room. For him it was simply some place else.

> [Plum]: It [the robot's physical environment] is nowhere to be seen, it somewhere else, no real distance, just somewhere else.
> [Question]: If you were to describe that somewhere else?
> [Plum]: It's just somewhere else, you know, I would not need to know where it is, it is just somewhere else, I could describe the looks of it but I would not need to know where it is.

Of course system breakdowns need to be minimized and correspondences need to be within certain limits. For the most part the robot seemed to be within these limits. The only events that caused confusion were large differences between virtual robot and physical robot positions and breakdowns. These events clearly affects ability to focus on the task and as a result presence. Correlation between the spaces is important if the video of the remote scene is shown. Correlation between the virtual robot agent's position and the physical robot is crucial for viewing the monitor metaphor and for placing the Reality portal textures.

Lag and Error Handling

The issue of lag and robot breakdown emerged as a separate issue. This lag was realized in different forms. One form was a lag between the movement of the virtual robot agent and the physical robot when the video was displayed. This lag is estimated about 500-1000ms, which is more than enough to notice. Another form of lag was the time before the effects of an issued robot command could be seen, *i.e.* the time before the robot could be seen to move. Breakdowns, when a system component failed and needed to be restarted were not directly visible. They were only noticed by a prolonged wait for a response. Thus occasionally, when there was significant lag, it was not clear if the system had broken down, or was just being slow.

As Plum and Orange point out, this can make a user unsure of what is happening:

> [Plum]: It was not as direct as you wanted it to be. The delays kind made it like you told it something and you had to wait a while to see if it did it and then you had to try it again.
>
> [Orange]: There was a point where before the program crashed, it got stuck in the corner of the room, I didn't feel sorry for it, I was irritated.. I am pushing the right buttons and it is not moving where I want it to move. It is stuck there and I think I am doing what I am supposed to do, so what's the problem?

This clearly can be a point of frustration, destroying the contract that the user enters into when using the interface. Some feedback needs to be issued to let the user know their commands have been registered and to give some indication of what the problem might be.

> [Yellow]: I think you should get some feedback so you understand why it doesn't work, it just didn't operate and I was not sure if it was a situation where it broke completely or I just couldn't do it. So you should probably get an indication for each of those two situations, or at least if it is a normal situation.
>
> [Mauve]: I think it is important that the user gets some feedback from the robot. The robot could go to the wrong room and you may sit for 10 hours.

With this system there are at least three layers where feedback could be given. One of these is the point of issuing the command, the act of clicking on the arrow for example. The next is when the virtual robot agent receives the command, and the third is when the physical robot receives the command. Different decisions for each of these can be taken. During the pilot runs in the first day of the study, it was realized that there was no feedback on the graphical arrows. After those first sessions, a behavior was added to make the arrows

blink when a user clicked on them. This was a good solution to feedback at the first layer, and resulted in far few repeated frustrated clicks by subsequent study participants. It is felt that this may be enough feedback for the user to know the command has been issued. Blinking the robot one second later when the command had been received and a confirmation returned could be more disturbing and make the presence of the communication link unnecessary opaque.

However there is a need to indicate to the user when the robot is having difficulty. In some instances the robot could not perform a requested navigational command because there were objects blocking its path. In these instances the robot simply did nothing. Many users guessed from this behavior and its proximity to an obstacle that this might be the case, but it requires a user to hazard a guess. When asked about how these situations might be communicated, Magenta had a suggestion:

> [Question]: How should it communicate [its] capabilities?
> [Magenta]: It should show without having a sign or text. It is difficult. It could have some lamps, like a traffic light, red yellow green, that is very simple..

Using the visual channel like this could be very effective. At a minimum the robot could send a signal to the virtual robot agent that it cannot move because the forward path is blocked. The virtual robot agent could then change the color of the virtual robot model to indicate this state. Changing from the normal color of blue to red is probably an intuitive signal that something is wrong.

One should also be careful not to overload the amount of feedback, as Peach says, sometimes a user may just want to issue a command and assume it it will occur unless otherwise indicated: "I don't want to know what is happening along the way, just go here."

Human-Robot Relationship

This section contains responses about the relationship of the human and the robot during the activity of using the system and performing the task. How does a human user consider the robot? What are the different roles the robot might take? What subjective view does the human report to have? Do they feel responsible for it?

I am the robot While performing the task, a number of respondents reported that they identified with with being the robot.

> [Azure]: Like for instance, in the robot's view I sort of feel like I am the robot. So I feel like I am moving when the camera moves.

> [Khaki]: I tried to identify with the robot, yes consciously I think I did.
> [Question]: Did you then think of yourself as the robot?
> [Khaki]: Didn't feel I was the robot, but I wanted to have a view from the robots point of view. Yes I was the robot as I was at that position, which may be a [only] semantic difference.
>
> [Orange]: I did what I would do if I were the robot. [...] As the robot, yeah, so I guess I did have some sense of being there as this thing.

This was especially true when looking from the robots' view and moving with the robot. While in that view the user's avatar is attached to the robot's avatar and when the virtual robot agent moves the user's perspective changes accordingly.

I am not the robot For others the robot was thing, a tool. Most of the people in this category, reported that it was a matter of control. It was not them that were acting, but they were controlling an 'other' which was under their command.

> [Question]: Did you feel your were the robot?
> [Mauve]: No I didn't not at all, I was steering something.
>
> [Question]: What relationship with the robot?
> [Maroon]: Controlling. I have been working too much [with robots] to feel emotions.
>
> [Green]: I cant say I am in it, but I can say that I am in control, so it depends on how you define it. I don't think I am ever going to get the sense that it is me.
>
> [Yellow]: It really obeyed me, it did not make any moves that surprised me or make any movements I did not expect. It was easy to control, the soft movements worked well with the arrows, so ah, I felt quite friendly to it in that way. It was nice, and not hostile anyway, it did not try to run me over [chuckles].

However making this connection of control or identification is surely dependent on the quality of communication with the system. As stated, a number of times the system broke down and this affected the relationship the respondent had with the system. Lime reports that the relationship changed when the robot system stopped working: "When the robot was working it was me, when it wasn't it became something else, less personal." What is noteworthy is that the user can slip in and out of that role as well as manage the separation when it fails.

Silver had a unique point of view on the system. His is the view that the CVE offers a view inside the robot's "cognition," a view of what the robot knows about.

> [Silver]: But the point of this thing is that you have the system's point of view in the world. I think of this as the robot's cognition or view of the world. It helps for my thinking to know the computer [robot] is looking at those things right there. The robot has seen these flags, and that helps.

Caring for the robot Respondents had different responses toward caring for the robot when it appeared to be in trouble. Although the robot was able to competently handle collision avoidance in the environment, there was one location that was more problematic than the rest. This was navigation around the shelves that stick into the room. The robot never collided with this piece of furniture, but often took more time to navigate when this obstacle was in the path. From certain viewpoints, the robot might have looked like it was *about* to collide with the shelves. The reaction to this event was individual. While some respondents pushed or were ready to push the stop command, others leaned back and chuckled.

> [Lavender]: When it went around that corner I thought it might bump into a corner, but it didn't so I figured it had some kind of knowledge.
>
> [Mauve]: I was really worried when it went outside its boundaries, I thought it had gone astray, but I guess you have some kind of obstacles avoidance that pushes it away from things?
>
> [Lime]: I cared about the robot hitting things.. I did not want to knock the shelves over. That may be a boy girl thing.
>
> [Question]: Did you rely on it [the robot] that it would stop? [Orange]: No, I had reliance on me that I would stop it.

Purple was happy to sit back and see what happens: "To me its a video game [...] that's its problem." As Lime points out some of these differences may be gender related. That discussion is beyond the scope of this study, but it can be said that in both *caring for the robot* and I am the robot the splits are almost perfectly along gender lines. This division would suggest a further investigation into these phenomena in order to see how it might affect design decisions.

Changes with video on When the video of the physical environment was turned on (the camera view), the perception of the robot changed somewhat. This change in perception was completely different for Magenta and Purple.

For Magenta the relationship became stronger when she could see the real robot. For Purple, when he could see the real robot he slipped out of being the robot and stepped back a bit.

> [Magenta]: Yes definitely, when you see this [virtual robot agent] you don't have a relationship to the avatar so much, when you think I am running around a thing that does not have a cord or anything and you think it is just me running it, you get some relationship to the machine. It is like when you talk to somebody who is an animal or a little kid which doesn't communicate, yet you can still get some sort of feedback, like when a baby grabs your finger, some small communication.
>
> [Purple]: Before you switched this one on [video] it was me going around, like in a computer game. It was me going around. I saw myself standing behind the robot or whatever sitting on the back of the robot- like the computer games.

This is probably based partly on an aesthetic predisposition. During the interview Magenta pointed out the failing of the CVE graphics and how she found the CVE unattractive. However Purple repeatedly indicated his interest in 3D computer games and how he saw the robot system as another such game. This is all to say that predispositions play a role and for some people a deliberate effort may have to be made to overcome these.

Magenta was also had her guard up regarding feelings for the robot. She responded that she was weary of feeling something for the robot and felt this might be easy to do.

> [Magenta]: I am just thinking that the problems I get when talking about the robot, you get some feelings for it, like it is entity or thing, I almost asked you before if the robot had a name. It feels like if I am the user, than I will develop some relationship with it, it is my friend and it doesn't like that, blah blah blah. Perhaps it is then easy to mix up humanoid activities with what you want to do. Or that you want the robot to do as many thing as I would.

This again points to care that should be taken when designing anthropomorphic features. Perhaps this is something that can be configured by the end users, for instance in a domestic robot application. Offering the end user the possibility to select features may change the relationship to one of control, making the user less weary about such emotions.

Remote manipulation Two of the respondents reported that their feeling of presence would have been stronger had they been able to change something in the environment. That in the role of *observer* they did not feel presence in the environment, as much as much as viewing it through the robot:

> [Lavender]: It was only a video camera and I think that is because of the task, it didn't have any social function. If it suggests things, and some things are automated, then you get some added value than only a way to perceive the environment. Perhaps if it grabbed things, or got things to you, or did anything except [just] moving around looking.
>
> [Maroon]: I was thinking before that there was this question about "how did you feel the presence". You would feel it much more if you could do some manipulation. You would feel much more like being there. Ok I am moving the robot, it is an observer, and that I am not really present.

It is unfortunate that the arm system was broken at the time of the study. To investigate this sense of presence it would be worth performing further investigations of manipulating and changing versus observing a remote environment.

Division of Labor

When did the robot and the respondent establish a working relationship on the division of labor? This division appeared to be flexible and center around the robot's perceived competence.

There were two methods for control, one for fine control, and one for specifying the end point to navigate to. While the fine control offered local adjustment the point-to-go interface provided a simple command to give a final destination. Mauve points out: "Here you let the robot walk around, and that is very nice not to think about, you just click and it goes there." However with point-to-go interface the robot employed collision avoidance and occasionally was seen to be having problems. Though the robot was never in real danger it tends not to take the same path a human would around an object. This is because it makes local decisions and sometimes will take the wrong way around a bookcase and then must backtrack and try the other direction.

Division of labor The robot and the human are collaborating to solve a task. The level of this collaboration shifts depending on the immediate activity. At times this is at the level of "go to that box" at other times this reduces to the level of "getting around an obstacle." The system is designed to enable this shift in collaboration level to happen implicitly based on actions that are made. In the instance of path planning, this level shift happens interactively as a user helps the robot with local obstacle avoidance. The following respondents report on their choices for helping the robot:

> [Magenta]: It is a certain delay before it starts when you click on the floor and over here it seems to get stuck on the shelf, 'zwip' it tried to walk through the shelf, so I clicked out here so it would

> walk out here and then go there. I don't know if it could have understood to walk around it.
>
> [Mauve]: I tried to help it by giving it a more straight way away from the bookshelf, yeah partial path-planning, just to say that "ok it is difficult for you to do this, so go this way instead."

At some point these respondents made the decision to help the robot. The level at which this happens and the amount of time before someone decides to intervene varies between individuals. That is to say that the specific circumstances when someone decides to offer help or takes control is individual. This implies that any design decision that tries to force this circumstance at a certain level or time will be inappropriate in some circumstances. For instance Peach was prepared to take control but decided not to, even though the circumstances for doing so were almost the same as for Magenta and Mauve:

> [Peach]: If there was no movement I would have helped, I would have probably have backed out the other way. Then I would have chosen the other, I mean if you can't go like straight like this, I might have to choose the other and go around.

The same individual may make a different choice of when to intervene depending on what has happened before or what they expect to happen. These variations reject the notion of a strict division of labor and make a case for a *working* division as discussed earlier in this thesis. The user may also have different reasons for helping or intervening with the robot behavior. For some of respondents this was to help the robot, for Maroon it was to improve the speed of the task:

> [Maroon]: I gave more direct commands, because I had the feeling it would go faster. First I tried the direct command, but when I give partial tasks it will go faster.

Thus not only are there different levels and times involved, but different motivations. For Orange all of her decisions were based on her model of the robot. This implies that even though great care needs to be taken in the presentation of the interface, perception of the robot capabilities will depend on experience built with this robot and with robots in the past.

> [Orange]: I felt like I completely adapted my behavior for the robot. I mean I did everything based on what possibilities, of its limited capabilities.

Feedback from robot Some respondents asked for more explicit display of the robot capabilities. These might take the form of graphical representations of robot intentions.

> [Mauve]: I got a view that you should look at the robot and the user as an entity that tries to solve a problem and I think it would be very important that the robot knows what it can do and that the user knows what he can do. Otherwise for instance planning would be very difficult.
>
> [Maroon]: So for example, when there is no path planning and I click on a point, there could be an artificial line showing that "oh I just go straight" or "I go on paths around it" which should tell me as a user "oh I should be more helpful," or [that] it is doing fine.

As Silver points out in the discussion on Relationship, the CVE can act as a repository of information for which the robot knows about. This can be seen as a user interface concept where the presentation is used to make explicit the knowledge of the robot. This was not strictly true in this system as the path planning system was not made aware of all the furnishings in the room. A user might assume that because furnishings are displayed in the CVE that robot navigation can take advantage of knowing their positions beforehand. As stated however the current system employs local navigation. A navigation sub-system that is based on using more of the information in the virtual world is an obvious extension.

Endowment Respondents were purposely not told beforehand about any robot competence. Given the participant backgrounds in the study there were surely respondents that brought along expectations based on their experience of previous robots. However the point made here is that this understanding of robot competence is also something that is built up by working with the robot. The following are some typical comments from respondents:

> [Mauve]: I helped it a bit, but that is because I have worked with similar robots for a long time and I know that it seemed like it was trying to get right through the bookshelves. And I know you have some algorithms to take it around but it takes it a while for it to actually do that.
>
> [Plum]: I could see that it avoided some obstacles, and it stopped, it had some sensing or preprogrammed model and started going around the shelf for instance. I did not think it had any [computer vision] or was analyzing the 2D video or anything.
>
> [Magenta]: I think you have some algorithms that say it cannot go in certain areas. I did not think it knew it could go around the shelves. [Magenta tries this action] OK I see.. but still some problems.

During the interviews it was often possible for respondents to experiment with the robot to try out features they wanted to understand better. During the

task session Magenta intervened by stopping the robot and issuing a partial path to the robot in order to get it away from the bookcase. Magenta was asked to try moving the robot around the shelf during the interview. As a result of trying this with more confidence that robot would not hit the bookshelf, Magenta changed her perceptions by trying to let the robot move around by itself. It was clear this still bothered her and would take some time to fully trust it. This again points to an interface that enables individuals to intervene as they see appropriate not by interface design. It also indicates that a person's understanding of robot competence is an active process they undergo by working with the robot and working with the best way to convey this competence to a user.

Trust

Trust arose as an issue of concern for some respondents. This section centers on that issue and speculates on some of the factors based on what respondents have said.

One element of trust is the video representation versus the representations from the virtual world. Occasionally inconsistencies between these two views could cause confusion. Here Azure makes a statement about these different aspects of the views:

> [Azure]: The video on the top of the robot? I think I can trust it, the video, since we are all used to video. But the virtual, I don't see why I should trust it. I feel like I can trust that picture [video] because it is like I am seeing it.

Purple became confused when the video switched on. He had become absorbed in the virtual world and found the video of the real robot confusing:

> [Purple]: Yeah I was a bit confused with the video and the VR.. there was an inconsistency. [...] I was not sure, I was thinking which was right, but it must be this one [video].

Surely also some of these perceptions are influenced by previous experience with the medium. But it is also clear that with a virtual environment almost anything can be created and there is no guarantee it represents something from the real world and that video would be more resistant to such a false construction. It would then seem intuitive that more trust, vis-à-vis faithfulness to the real world, would be given to the video.

The following quotes indicate that the presentation of the system may be a delicate thing, as is a common belief in research and this is certainly true in folk aphorisms about "first impressions."

> [Maroon]: Trust of a machine is not such an easy thing to grasp. Users tend to trust machines until the first problem, and when they

> realize it is not them and the machine, then there is more or less trust in machines, and it also depends on your education, what you study.
>
> [Silver]: You will always get into the situation where something you want to do is not built into the system and you will want to control the robot by hand, and you need the system to tell you 'no I failed'.

This indicates that there is a need for honesty in the presentation of the machine and its capabilities. This relates to the earlier discussion (Chapter 2) of attribution (the qualities the robot has) versus endowment (the qualities the user associates with the robot). If there is a large gap and a new realization is made this gap may be received as trigger for broken trust.

Magenta shared some concerns about it being a good thing that the robot does not have a personality:

> [Magenta]: I think it is a good thing it is just black nothing, as soon as you put smiley face or eyes or arms, than you take steps to make it more human. It could be good to have feelings for the robot or avatar, but you have to try it out to see if it is good or not.

However this is a difficult balance, and there is some ambivalence in Magenta's statement. As stated, research has shown that people cannot help but anthropomorphize objects. This responsibility then falls on the designer of the system. Such decisions about anthropomorphization also depend on the application. For a robotic pet or toy, such features may be the sole purpose of the robot, for a machine that is to be trusted in daily household duties, such features should be treated with care.

Avatars

Avatars were not brought up explicitly as a concept in the study or the interface and the respondents were not given specific instructions on how to manipulate their avatars. However a number of respondents were familiar with the concept before. Some of these respondents questioned the idea of multiple representations for the user and robot.

> [Khaki]: ... it seems a bit bizarre to me to have this separate avatar. If you wanted to strengthen the feeling of presence you wouldn't want to make the separation between the robot and oneself, the robot is your avatar and that is all there is. But now it is more like there's a virtual copy of me that is running this other remote object which is one indirection too many I would say.
>
> [Magenta]: I think it is difficult to to know 'what am I?' in the virtual world. I have the impression that I am an avatar too in this world. It would be interesting to actually be the robot.

> [Khaki]: You could say give me an overview, the back view, top view, video view, etc., but only one body [...] While the out of body experience might help for checks of different kinds, but then you become present somewhere else. So maybe you could say, I have a body here and clairvoyance somewhere else.

These quotes bring up a notion that questions the standard notion of a user and avatar in a CVE. Currently the user and the robot have avatars in the CVE. If a third user were to come into the environment they would also have an avatar and this representation would be visible to other users in the CVE. What the respondents above seem to be asking for is a way to become the robot entirely. This is something that has yet to be experimented with. The primary obstacle to this is how to present the video. However with a good working RP system or with the proposed video screen at the end of the robot view cone, unifying the user and the robot might be possible without losing any great functionality. The user takes the view from the robot's exact position and then can navigate large space via the map interface. This confusion about avatars may have been underlying some of the confusion about the placement of the monitor metaphor on the robot.

Speech

The speech system that has been used together with the robot system and described in the previous chapter was unavailable for the study. However the idea of whether speech would be appropriate came up with a number respondents. For the most part this came up with HCI researchers who had some experience with speech based systems. Both speech input and speech output were considered.

A few of the respondents made the observation that speech was not necessary for this task. That the navigational command interfaces provided were sufficient for the task at hand:

> [Yellow]: Of course it would be nice to be able speak to it, but it is not necessary to have that control.
>
> [Silver]: Well for this task, it is not obvious you need it. Linguistic interaction is not crucial here, it helps with more difficult tasks. What you do want to do here is perform an action and save that action to be repeated again. That would be more important to me than being to speak to the robot.

What Silver asks for is something that might require speech. Without developing a special signalling system, it would be hard to communicate by graphics alone that the robot was aware of this task and could repeat it. Lavender asks for a similar feature using speech:

> [Lavender]: No, well maybe, I would have liked it to do the task for me, and then speak, and say "ok I have done this task before I can do it know, do you want me to" and just show me around.

As soon as competences such as these become available it might become necessary to introduce speech. Speech would also be preferable to a menu system as the speech would preserve the graphics interface. Text in 3D virtual environments because of text's 2D nature is often problematic.

To speech or not to speech? Having a speech system opens up another modality for interaction. It might be preferable to other simpler forms of communication, but this new new modality needs to be designed and maintained.

> [Yellow]: Yes of course, voice is probably better than a beep. If you have voice and use also voice output that would be nice. But that means you get another channel, because you don't use voice anyplace else, that may be overdoing it a bit, but why not?

The perception of the interface changes with the addition of voice. As Magenta points out, speech is a "loaded" area that has a number of parameters that need to be worked out by careful design.

> [Question]: Would it be different if it spoke?
> [Magenta]: Then you would develop more relationship with the robot. I know when they did some testing in the [my lab] I felt like I wanted to have communication with the robot. [...] When you start developing an identity its really political stuff you get into, how should it speak, should it be rude, nice, should it squeak? If you want to prevent people to get close to the robot, I think you should keep it very machine like.

This question about speech relates closely to the other design questions regarding anthropomorphization (*e.g.* adding eyes, smiles, etc.), in that it brings with it a number of cultural and personal preconceptions that might be endowed onto the robot. This may change the user's expectations in unpredictable ways and open up the opportunity for an emotional connection that might be completely inappropriate for the task.

Click on Box

Many design suggestions came up during the study. These are listed in the next chapter. A number of these design suggestions were implemented in the time between task sessions with respondents. However one such suggestion required extra functionality that could not be created in-between sessions. This was the idea of clicking on a box to go there and look at the box in one motion. In

the provided interface, a user had to go to a location first, and then rotate the robot to see it. This idea was taken up in the interviews.

The functionality of this proposed feature was not hard to convey to people that had used the system for the task. The functionality is succinctly by Khaki: "If I click on this box, then it means translate me to this particular object and look at it."

The interface is designed to be able to support greater competence on the part of the robot. In the current system there is clearly a need a need for higher level commands, as Silver points out:

> [Silver]: What you want to be able to do is record macros, 'this is what I mean by checking the mailbox' and then have the command 'go check the mail box'.

A number of respondents found it problematic that they would navigate to the box and then have to turn the robot manually. Mauve represents this view, and it was this suggestion that led to exploring this idea as the next feature to implement:

> [Question]: What about the robot not turning toward box?
> [Mauve]: That was a bit annoying, you would like the robot to go to the box and look at it.

Peach and Silver testify to the usefulness of such a feature and Silver points out that the addition may not be a solution for all users. As some would still have difficulty using the overview map.

> [Peach]: Because when I went here I had to adjust it to look at a certain object. So you can say not only walk here, but grab this object "go here and grab this object." Yes that would help.
>
> [Silver]: Yes, [clicking on boxes] would work absolutely, because I have this representation here and I like it and of course I have used 3D games and I have played around with maps, but I know there are people that cant click on maps, because they wouldn't. They would much prefer this view (where they share the view with the computer), and then clicking might be less useful, as clicking there is not much better than driving the robot there.

With a redesign of the monitor view on the robot as proposed in the section on views, it would be possible for a user to click on boxes they see in front of the robot. Clearly with such an interface the manual feature of rotating and translating the robot would not disappear. The box-click navigation feature would be another tool to use as appropriate. Although this feature has not been tried out with users it is under development.

Applications for system

Respondents were asked about potential applications. One particular question was if they knew of tasks in the home where the robot might be useful. A number these suggestions are offered below along with some speculations on how these might take form.

When asked how such a robot might help them in their daily domestic or work activities:

> [Green]: I know people that have a need for it. But I don't know if I have a need for it. Maybe when I am at work and I need to turn on the TV at home, like a video.
>
> [Azure]: in my work in biology, we use robots, like when you have to mix something very carefully and of course, I haven't worked with radioactive material, but then it would useful.
>
> [Orange]: If I looked at this room and said, "oh yeah I want that book", I want it to get that book and I could immediately click up to a sort of floor plan view, which would show me where I want to go and where the robot is, I could picture that.
>
> [Silver]: I would love to have it in my home, a machine like that in my home. ... Yeah, if it wasn't obtrusive and the batteries were cheap, etc. I could have one of these running around my home. What I do need is something to look at and for stuff at home. Something to read me a piece of paper. I forget. I have this notebook, with little pieces of paper. I don't use a palm pilot anymore. I need something to open up [folders] and go through them. For example, "lets go check the bookcase" ah that is what is in the bookcase, and then I want to say "that book, click, now fax it to me" and the next time I don't want to go through the hassle of navigating through the book shelf and just say "now go to the bookcase."
>
> [Azure]: I think it would be useful to be able to program the robot to do a series of things. Like back to biology, where for instance you have to do things at completely equal amounts of time you could program the robot to do things at exactly, like 5 minutes, which is hard work for yourself to do it. But then you have to completely control the robot of course, otherwise you don't know what you end up with. [...] Hopefully a series of experiments will be completely equal, but with the human factor you never know if they will be the same.

Others found no real use for a robot in their home. Peach felt that most automated tasks could be done better by specific appliances for those tasks.

Lavender felt such a robot, if it had a camera like the one in the study, could be an invasion of privacy.

> [Question]: Would you find the robot useful for searching your home?
> [Peach]: Not me, no I have never had, and I am not paranoid in that sense, wondering if I have turned out the heat or something.
> [Question]: Or checking if that letter from your girl-friend has arrived?
> [Peach] Maybe, I would like to know for instance, what is in my fridge. I am having a dinner party, I want to know whether I have pepper, but then maybe I would have an intelligent fridge instead of a robot. as I understand this BlueTooth technology, you might actually direct your machines remotely, then you don't actually need a moving thing in the apartment. But then again, well, but then, if you have a video cassette in your home, and you want to see a clip, you need to fetch the tape from the shelf and put it in. But there are not many applications where you need that kind of thing.
>
> [Lavender]: Maybe, not in my home ... because I know my home, and I actually wouldn't want one in my home if it could look at me. I wouldn't like to have it in my home, because then the norm could be "don't you have your robot on, [you] boring person"? [...] It does not disturb me that something is moving around, it is that someone may observe me from somewhere else. But I don't think I would buy one, because I don't see the added value. But I would see added value if I could go underwater, or underground cave, or like wandering around mars, or into my own veins checking cholesterol or something.

The last comment opens up an application for such robots as a shared resource for the community. Robots in dirty, dangerous, distant (DDD) applications have traditionally been used for the purpose of research or exploration. Another application might be to have such robots available as entertainment or edification resources to explore those places where general citizens are prevented to go due to cost and other barriers. An example of such an application was the brief and *ad hoc* WWW access to the Mars Rover virtual maps as they were being discovered. This was perhaps the most popular tele-robot to date, and more such applications, inter-planetary and terrestrial can be easily imagined. It is possible that as the capabilities for broadband interactive media become more common such applications might become common.

> [Orange]: ... what you want to do is say "go find Tolstoy's war and peace and bring it to me" but I think we are a little bit far away from the days where we can say that.

Orange's request is not that unrealistic. Using some of the techniques presented in this thesis and limiting the scope of the search it might be possible to construct just such a system.

5.4 Final Comments

All the respondents gave the impression that they enjoyed working with the system. They also gave the impression that they appreciated being offered the chance to work with the system and offer their comments. In fact every attempt was made to show appreciation for their candid comments. Below are some of the final comments from the respondents, giving a feeling for the mood of the sessions:

> [Silver]: I want one of these.
>
> [Khaki]: It is fun, it is actually a bit more responsive then I thought, it moves faster.
>
> [Magenta]: I don't think the video was on, but it would have been interesting to see the shelf falling because the robot was hitting it.
>
> [Plum]: It was fun when it runs.
>
> [Lime]: It was fun while it is working.
>
> [Purple]: Probably if my mother doing this task, I mean she would be completely confused, just by pressing buttons and realizing you could switch views.

Conclusion This chapter has examined the data from which a number of design choices and suggestions have been made. By reading this chapter it is intended that the reader has gained a deeper sense of the system, the study, and an increased understanding of how the people related to this system and some of the design implications that have been made. In the next chapter the findings from this study are brought together. In that chapter some of the analysis made here is revisited, summarized and made more concrete.

Chapter 6

Study Findings

What is offered by the previous chapter is a chance to "get inside the heads" of persons that have used the system. This first-hand presentation of using the system offers an understanding of the system that could not be gained by a simple textual description of the system. The intention is that it is enriching reading and the understandings gained are not necessarily tangible. However this chapter makes concrete the lessons learned through the study by summarizing the main points.

To ground the generation of new ideas, identification of the needs and issues of the current implementaion provides a starting point. This identification can help focus attention on the problems to be solved. One of the purposes of performing the study is to locate the problems that can ground such new idea generation. Working with users of a system on a particular task within a study of use is one way to expose these problems and issues for further exploration. This then grounds further ideation and design iterations.

This chapter presents a summarization of the Study results and reflections in the following sections:

Study Discussion: A reflective discussion about the Study itself, how it was run, what was learned, what future studies might be done.

Design Suggestions: These are the concrete design suggestions that were made by participants in the study.

Development during study: During the study the system was changed to respond to participant feedback, these changes are presented here.

Study design implications : While design suggestions are direct comments for improving the system, design implications come from the listening to what was said. These are a number of the issues and problems that were identified in the report of the study from the respondents. Many of these

have implications on future designs and these issues, along with concrete future design ideas, are presented here.

6.1 Study Discussion

Below are a number of brief sub-topics that center on the nature and experience of the study.

Complex system and complex environment All the participants seemed to understand the task, the robot controls and the robots capabilities (though one was confused initially). Few questions or suggestions from participants centered on any fundamental confusion about the set-up. Although some of the participants were aware of the system from before, others were completely from the outside. The fact that the system seemed to be accepted by all participants was rather surprising as a robot system such as this is not common nor simple. The fact that such a complex system was made accessible to persons that have never used such a system before speaks for the ease of the interface. It was also a complex environment that contained a great deal of information. Many participants likened it to a 3D video game. That these games are well part of contemporary culture also must have had a significant influence on the understanding of the interface and system.

Task-based design The users were given a concrete task to perform. This helped to ground the purpose of being in the situation of controlling a robot. The task was such, however as to elicit a certain type of behavior from the participants. Since the task was defined in detail, the participants were focused on finding flags in the environment and were quite goal directed. Had the task been more "explorative", it is probable that a completely different behavior would have occurred. For example, users were asked to search for flags in a certain order, therefore they filtered out other information that was in the category of "not a flag." Had the users been asked to perform a task that was more explorative e.g. "There are three things you need to find in this world, you will know them when you find them." There may have been a completely different pattern of filtering and behavior and as a result: different design responses. This is to say that the task request must certainly be taken into account when evaluating the behavior and comments reported in the the study. For a different style of task, some of the design suggestions might be different. This is in fact one of the implications of the study, that the selection of tools, views, CVE-fidelity are task based and may need to be user-configurable.

Polite participants Participants were surprisingly forgiving regarding system crashes. Much of this is due to many of the participants being from a research background. Those that were not seemed to accept the system as

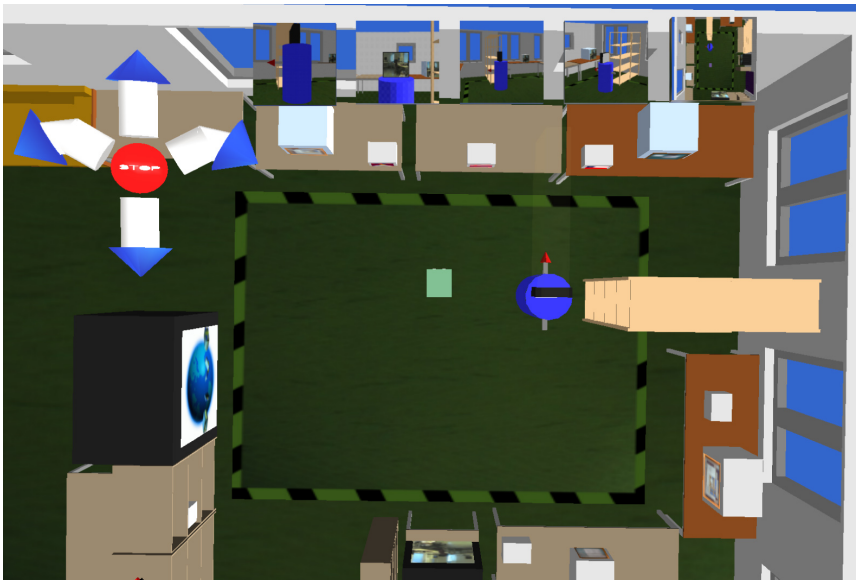

Figure 6.1: Top view. The top view looking down onto the robot and the environment was mentioned by participants to be the most useful (see also color figure **F**).

"experimental." Still it is not fully obvious why there were no complaints or expressions of frustration. As mentioned in the previous chapter, there was a strong impression that the participants appreciated being part of the study and that they enjoyed what they considered to be a unique experience: controlling a robot remotely. This behavior was probably a mix of politeness, acceptance of a complex system, and that a number of persons have become tolerant to "demo-disease" (that sickness systems get before a demo).

The squeaky wheel It is no surprise that there is more to say about the things that do not work as opposed to the things that did work. It was hard, for example, for respondents to speak about the reality portal flags in the environment. They were the focus of the task, and it would have been hard to perform the task without them. The other reality portals, not part of the task, however, were "clutter." This has led to one conclusion that the need for these displays is directly related to the task. To find out more information about the use of reality portals in a CVE a more complex study with multiple tasks might be run to discover more about a design that involved reality portals both when they were part of a task and when they were not. Also the interface of clicking on the floor produced little feedback other than "that worked" or "it was good to teleport to places." The click-arrows interface, in contrast, received

a significant amount commentary about it being "annoying," or tedious. Bad system design is perhaps much easier to identify than good design.

More experience Though some were familiar with the work, none of the participants had ever run the system before. During the introduction, the participants were introduced to the controls and then were given instructions for the task. It is possible that the deeper comments of expertise might have begun to emerge had the participants been able to go through the task again. The resource requirement here would have been too high on both sides. The study as it was, asked for over an hour of the participants time, and generally two hours of the person running the study in order to be sure it was set up for the next person. This would be something to consider for a future study, to run the study with half the persons, selecting over a broad spectrum, and increasing the length of the session.

Iterative development During the study, the system was changed to respond to comments from the participants and from features noticed by myself as system designer. This then represents not a "controlled" study in the sense of an experiment, but a design study in the sense of iterative improvement for the sake of building a better system. It was through this process that a partnership with users could be obtained and a number of features could be investigated. These changes are highlighted and discussed in greater detail below.

Successful Study Overall the study is considered a success. A number of problems with the system were found, a number of additional features and fixes were made and a number of new issues have been identified and made available for future exploration. The enthusiasm of the participants is also a side-benefit that should not be undervalued. Working with users not only brings questions and demands to a system, it renews and awakens interest in that system. On top of that the identification of issues not easily solved offers a base for the development of new solutions. Thus, at a minimum a study such as this results in:

- A more robust system;
- New implemented features;
- Renewed enthusiasm;
- Suggestions for new features;
- Identification of important issues;
- A platform for additional brainstorming.

The next section will discuss those concrete design suggestions given by respondents during the study.

6.2 Design Suggestions

One purpose of performing the study was to elicit design suggestions. Such design suggestions can be found embedded throughout the previous chapter's discussion with the respondents. This section, however, contains those concrete design suggestions made by respondents about the system. A number of these were then implemented between sessions if time permitted in order to receive feedback from other participants while the study progressed. These suggestions are presented categorized by the part of the system for which they are relevant.

Suggestions about robot camera control

> [Orange]: It would be nice if I could tell it, come in front of this box facing the flag.
> [Silver]: For this specific task, I would like to pull the focus around and say "look here."
> [Mauve]: You would like the robot to go to the box and look at it.
> [Lime]: I would like to see the robot's view when in the top view.
> [Lavender]: You can't adjust the angle of the camera when you point on the floor.
> [Magenta]: I thought that clicking on the boxes would center the robot to be straight in front of it.

The quotes above point to a primary failing in the system, that it was tedious to move the robot manually around to look at a box. This was a feature that would clearly be appropriate for this task. Although a user could command the robot to navigate across the room from the top view they often switched to the behind-robot view in order to move the robot to view the flags. This was the motivation for the "box-click" feature discussed in the previous chapter. This is an excellent feature for the system, but unfortunately it required a new structure to be implemented in the base-level of the robot and could not be done in-between sessions of the study.

Suggestions on control

> [Plum]: Since there was a lag in navigation arrows, it would be nice to be able to see when the command has been executed, when you press it it could be like a button.

This suggestion led to an immediate implementation of a graphical blink when a user pushes on the buttons to give feedback that the system has received the

button click. The other change that was implemented was a red-dot appearing for a little less than a second when a user clicks on the floor for the click-to-go command. Both of these were implemented before the sessions with participants the next day. Once realized this was clearly a necessary feature and caused much less confusion for subsequent users.

> [Khaki]: When you just want to move in a given direction for a certain while it would be more convenient to just press on that button.

This suggestion led to an extension of the click-arrow interface. This was that if the user clicks on the body of the arrow, the robot would move an increment, but if the user clicks on the extremity of the arrow, the robot would go in continuous mode. Participants understood and appreciated this feature and it cause little confusion except when it was half-implemented when a participant said "I think it is a good idea to have step or continuous commands when you turn, but I would expect it to be the same with forward or backwards." This is because there had not been enough time between sessions to finish the implementation for forward and backward.

One participant suggested that it would be a good feature, in the behind-robot view, to "Click on the video to go forward." This was easy enough to implement, but was available too late to be tried by most participants.

Suggestions on higher-level control Both silver and green suggest the need for higher level control. That the robot should learn tasks and be able to repeat tasks it has done before. This is departure point for future exploration.

> [Green]: Maybe the robot could imitate exactly what a human does, if I did it once, the robot could do it a hundred times.
> [Silver]: I would like to give a command like "ok circle around the room and look for flags."

Feedback Magenta asks for some feedback when the robot is stuck or blocked by an obstacle. This is an issue for future exploration.

> [Magenta]: It could have some lamps, like a traffic light, red, yellow, green.

Camera control Peach suggests that a way to switch views automatically so the user would not have to choose views. Good work can be done here, but would require a structure that makes predictions on the most likely point of focus in the scene.

> [Peach]: You could also think about automatic camera control. When you move from one end of the space to the other you could have automatic camera control to follow the robot.

Video control Plum and Purple wonder if there is some way to have more camera control in order not to have to move the entire robot to look. This is in response to the robot being relatively slow in respect to the speed of a pan-tilt head or zoom and to the speed a human would pan in the real world.

> [Plum]: A way to turn the camera on the robot, so you could move the robot somewhere and pan around, instead of moving the whole robot.
>
> [Purple]: If I could zoom in I could check the flag without moving myself.

6.3 Development During Study

During the study several changes were made to the system. These changes represent iterative development to the system while working directly with users. These changes do not include the many changes that were made to the system in order to prepare for the study. For the most part, those changes are described in chapter four. Below are the changes that were implemented during the study, either in the time between sessions, or overnight before the next day.

- Move view icons: After the first informal pilot run, the view icons were repositioned to be at the top of the screen, which was much better than at the bottom where they interfered with floor interaction.
- Video on visor: After the second pilot, robot camera video was added to a larger button on the visor so that it is always visible as the user moves around. This is the same video stream as in the monitor metaphor that is attached to the robot.
- Shrink avatar: The user is represented by an avatar. In the top view, the user is looking down on the avatar and space from above. Although the avatar had been made small for this application, there was no reason for it to be visible at all, so it was scaled down to a minimum size.
- Click-arrow feedback: The arrow interface was changed to blink once for 500ms when it has registered a click interaction signal and sent the command data to the robot. This confirms, to the user, that the command has been sent to the robot via the virtual robot agent.

- Floor-click feedback: Similar feedback implemented for point-to-go interface. When a user clicks on the floor, a small red sphere appears on the spot for a fraction of a second confirming the command has been sent.

- Click-bug: A bug was found that crashed the virtual robot agent if the user clicked the arrow interface too many times. This problem was resolved by two fixes. The first is that the user clicked less with the feedback implemented above, and the second is multiple commands issued within approximately 500ms were thrown out.

- Continuous movement: The arrow interface was changed to add continuous movement in addition to simple increments. This was tried out first with turning and then the forward/backward movements were implemented. The user then had to use the stop button to stop when the end of the arrows were clicked. Collision avoidance is still active when using continuous movement.

- Adjust speeds: Initial speed settings for continuous movement and turning were found to be too fast for users. Different speeds were tested, but final solution involves setting the acceleration speed low and the velocity medium ($10CM/sec^2$ and $25CM/sec$ respectively). This has the correct effect of moving slowly for short distances, but speeding up over time for longer movements.

- Click-video: Some of the participants comments indicated that it would be natural for a user to be able to click on the video in the monitor metaphor to steer the robot. This was implemented so that if you click in the middle third the robot moves forward and on the side thirds it turns the robot.

- Click-box: Work was started to make the interface of clicking on a box to move to a predefined location in front of the box and then turn to face the box. A "kludgey" implementation is in place waiting for a more formal solution.

6.4 Study Implications on Design

The sections that follow summarize the more salient findings of the study.

6.4.1 Task-dependent system features

One finding of the study is that the appropriate detail, functionality or display of a number of features depends on the task at hand. These are for example:

- The fidelity of the CVE model. It was confirmed that a sparse model can be a benefit. The desired amount of detail displayed, or fidelity, will most likely be linked to the task at hand.

- The views available on the CVE. The view that a user has on the CVE will depend highly on the task. For this task the top-view and the behind-robot view were the most useful, while the side-views were not employed. For another task, with different structure, it might be the side-views are critical.

- Reality Portals displayed. The Reality Portals in the model that were linked to the task were crucial, however the others that might have contributed to "pictorial realism" were not seen as necessary. Thus the value of displaying a particular Reality Portal may change for a particular task. This had not been considered before, so an effort would need to be made to enable the control of Reality Portals from the user interface.

- Control methods. It was discovered that a new higher-level command, the box-click command, would be useful for this task. However the available commands such as this, that are simple composites of given commands, may change depending the task. This has implications on their implementation.

The above finding suggests an interface that is flexible with respect to the dimensions given above and that enables a user to control these features with respect to a given task.

6.4.2 CVE issues

Views The top and behind-robot views were used the most. For this task it would make sense to leave out at least one of the side views. The top-view affords a map-like interface while the behind-robot view offers that of a vehicle. The zooming from one camera position to another was far from perfect. If this is to be useful, more care can be put into the sub-system that handles the transformation from camera to camera and to take into account virtual obstacles along the way. Also further investigations into automatic camera control can be investigated. Many 3D video games, are able, through careful design, to perform sophisticated camera control that create a seemless and enhanced experience. Some of these techniques might prove useful for task-based environment visualization as well. In response to respondents comments an alternative robot-view would be to unify the users and robot's avatar. This would require a redesign of the monitor metaphor, but might afford a view that feels more like "being the robot" for those that choose that.

Navigation For many participants the top-view worked fine for spatial orientation. For others, where maps are not as effective, one might try a strategy of adding landmarks. If the environment lacks landmarks, it would be possible to add virtual beacons that help orient the user in the world. These beacons could be subtle such as ambient objects that fade to the background when not attended to. An example of such a feature might be coloring the walls differently depending on their orientation in the room. Or they could be more obvious. Thus one might be able to passively absorb that they are facing "virtual north" because they are looking at the blue wall.

6.4.3 Environment visualization issues

Reality Portals: The RPs in the system were the flags on the box, the CRT monitor screens and the display of textures on the shelves. It was hard to elicit comments on these as they were either seen as belonging to the scene or alien to it. The desire for and amount of these may be tightly linked to the task at hand. Those RPs that were poorly placed, such as the texture on the middle shelf that one could navigate behind were confusing. The resolution of these images needs to be treated with care and should be as consistent as possible.

Monitor metaphor: The monitor metaphor confused a number of participants. To some it was clear, to others it required some understanding to make sense, others found it frustrating. The placement of the monitor on top of the robot needed some explanation to some participants. Because of the current implementation and technical limitations related to distribution, the image quality of the video stream is low, described by one respondent as a "scrim effect." Although it runs at 15fps, which seemed to be fast enough to perceive it as a moving image, the quality of each frame suffered some effects from the color model and compression employed. This would indicate that improvement of the video inside the CVE is a priority task. An alternative model to placing the monitor is suggested. This is to place the video in front of the robot when a user takes the behind-robot view. Then the video might appear like a special lens on graphical world.

Camera Video: The two side screens contained camera video for half of the study. Sometimes this proved useful, other times not. The screen on the right displayed the robot's view and this image was better quality than the same stream inside the CVE. When given the option participants, however, prefered a single view. Often this view was used for confirmation, while using the monitor metaphor to navigate. A number of respondents point out the limitations of the video, that is is hard to gain a spatial sense of the environment, and the robot is slow to pan

around the room. Both of these arguments have been motivations for the Reality Portal implementation. Another drawback to the video was that the environment was low-contrast and it was hard to make out features. Although this can be adjusted somewhat, it is not possible to always alter a remote environment or adjust the camera parameters. This problem could be partly solved by a augmented reality or augmented virtuality solution.

6.4.4 Trust, Presence, feedback

Trust Trust came up as an issue for some respondents: What they trusted in the system, what led them to have less trust. Respondents had different responses to the video as to the CVE. A number of users commented on the "realism" of the video, but asked why they should trust a graphical world. Other uses were comfortable with the CVE as they would be a video game, that is they did not question the medium and went about the task. This may be influenced by aesthetic or cultural predispositions. Trust issues also were involved in the relationship with the robot. These were trust in its autonomy, which was eroded if the robot broke down, but not as much as was expected. Also trust in terms of representation of capabilities, there, some respondents indicated an uneasiness about the robot developing a personality.

Presence A number of respondents commented on the relationship between the real and the virtual both when it made sense and when it did not. For most of the study the correlation worked well and most comments were given only when the position and orientation of the robot and the virtual robot agent were not synchronized. This was a factor that could influence the reported sense of presence. Another factor was when the system failed. System failure was notice by an unusual lag in response, as the failure was usually caused by the communication link failing, not the graphics. Thus this was first noticed gradually by the participant. If this were a re-occurring problem, (*e.g* because reliable transmission was not always available), it might benefit the system by making this connection more explicit. For example when the robot is "on-line" the virtual robot agent might be displayed in a different color. When a response from the physical robot has not been received in a certain time period (*e.g.* a few seconds) this color could be changed to another color that indicates this lack of communication. One channel that could be used is transparency. When the robot is online, the robot avatar would appear solid, if this connection becomes unreliable the graphical robot fades. A few respondents felt the presence would be stronger if they could do more than observe the environment, that is manipulate it as well.

Feedback Another item for feedback is the robot state. When the robot's col-

lision avoidance system detected an obstacle it would stop the robot and prevent it from moving further. This confused participants. This condition was usually "figured out" because of the robot's relative position to an object in the CVE. However this state could easily be made more clear to a user by signal. One good method for this, suggested by a respondent would be to use the color of the robot body. Normally blue, the virtual robot agent's avatar could become red when the robot's movement is blocked by an obstacle. Caution should be made not to over-build the interface. For the most part it was felt in the study that the robot should "say" nothing if everything goes as planned. There are three layers to the robot system, the graphical interface, the virtual robot agent, and the physical robot. All of which could potentially give feedback.

6.4.5 Relationship and Division of Labor

Relationship Respondents expressed different views on "being the robot" versus "controlling the robot." These occurred at different times and were split along gender lines. Most respondents expressed they were the robot in the behind-robot view, but that also these feeling could be enhanced by a repositioning of the monitor screen if that were a desire. Another difference was in how much responsibility the respondent felt for the robot when it moved around, *e.g.* caring if it were to bump into obstacles. This also seemed to be divided along gender lines. The implication here is that, although much more investigation would need to be done to confirm this, different perceptions of the robot and interface may require different interface manifestations. The relationship sometimes changed when the camera view of the robot came on. Sometimes this increased the connections, other times it startled the participant.

Division of Labor Nearly all respondents reported that they adapted their behavior for the robot to some extent. One manifestation of this adaption is the partial path planning many users did when the robot navigated around an obstacle. This happened spontaneously by the user "clicking" to a new spot for the robot to move to. This division is individual in level, time, and motivation. That is to say that at what point it happens, at what time, and why are different for different users. The interface should be open to these differences by enabling intervention at any level or time possible. Collaboration as planned for, does not realize the nature of collaboration as it happens in the world. This implies that any design decision that tries to force this circumstance at a certain level or time will be inappropriate in some circumstances.

Model Making Some respondents viewed the CVE as a knowledge database. That is that the CVE contains the world the robot knows about. Although this is true to some extent in the current system, this connection

could be made stronger, or distinction could be made between those that are known and those that are not. In principle this is merely a matter of implementation. Also it was apparentl that different respondents had different conceptions of the robot and its capabilities. In particular about how well it could perform obstacle avoidance, sometimes this changed after further demonstrations. This indicates that a user's model of robot competence is an active concept, the result of an individual working with robot over time. Working with the system is the best way to do build this this understanding (*e.g.* "by doing").

6.4.6 Control issues

There were a few different ways to control the robot, these were primarily the point-to-go interface where the user clicks on the floor, and the click-arrow interface where the user employs graphical arrows to steer the robot.

click-arrow The click-arrow interface was good for small distance and fine-tuning the robot position. However it was tedious to some. In response a continuous movement feature was implemented. This interface also required visual focus that might be in contention for focus on where the robot is going.

point-to-go The floor interface was good for long distances. It "teleported" the robot to another place and was considered universally the preferred interface for crossing the room. The problem with this interface for this task was that it was hard to turn the robot from this interface. This interface worked best for those that were comfortable with maps. Although it could be used in any view, it was most easily used in the top view.

box-click The box-click interface is the next step. It came out of the study that having this feature would be a clear advantage for the user in this task. Instead of using the floor interface and then turning the robot to see the box, the user would simply click on the box. This could work in both the top-view and alternative behind-robot view (described above). A preliminary implementation of this has been hacked together.

6.4.7 Applications suggested

Respondents suggested a number of applications and some concerns for using such a robot in their daily lives.

Look for things The robot might be used from work to look around the house for lost articles, check for mail, get a reference from a book, etc. It might help with biology experiments where the substances are toxic or where the repeatability is crucial. Another application might be to have a generally

available robot that enables a home user for entertainment or education purposes, to explore a volcano, a mine, their bloodstream or space.

Concerns Some respondents did not see a need, they would instead prefer to have smart individual devices that would communicate. Another would not like a camera in the home, and saw that as an invasion of privacy.

6.5 Summary

The study prompted a number of design improvements during the study. It also called attention to a number of problematic issues and pointed the way to some longer range design improvements. The information and experience gained provides a solid base for future brainstorming and further iterative studies with users. The study also showed that a complex system such as this can be presented to relatively naive users (with no tele-robot experience) in a short amount of time and successfully perform a non-trivial remote task.

Chapter 7

Conclusion

This thesis explores the problem of how to provide supervisory control of a remote robot. The solution offered is a system for human-robot collaboration via the medium of a distributed shared collaborative virtual environment (CVE). The key contributions are: the supervisor-assistant framework via the virtual environment and robot semi-autonomy; a selection of methods for visualizing video within the virtual environment; and methods for human-robot interaction and task specification. In addition this thesis has taken an in depth look at how these features might be used together to perform a task. To accomplish that a design study was undertaken where the responses were collected and a number of actionable insights were developed. The study helped to generate a pathway of further work to guide this, and related systems, forward.

This chapter reviews some of the key elements of the thesis from the perspective of the experience of the study. In particular these are: examining the supervisory control framework between the human user and the robot; comparing visualization methods, exploring interactions and task specification methods, and what might be required to improve the system. These topics are separated into sections divided by contribution components. In each section the dominant themes are reviewed and future directions are identified.

7.1 Supervisor-Assistant Control Framework

The framework, or the guiding metaphor, for supervisor-assistant control includes the use of a CVE as the communication medium and the robot semi-autonomy. The CVE is a representation modeled after the remote space. Robot semi-autonomy is the competence of the robot to perform certain tasks. In addition to the competence itself, e.g. navigation, it must also provide mechanisms for interaction with the behaviors forming the competence.

7.1.1 CVE design

An issue that was confirmed to be important in the design study was the degree of fidelity of the virtual environment. How much of the real world is represented and displayed can be seen as a crucial factor in the usefulness of the system. There seems to be no simple rule that applies in all circumstances. For instance, more detail is not always better, nor is a simplified world more clearly understood. The right fidelity, or model representation, depends on the task at hand and may in fact be variable. This implies a very dynamic world that is constantly shifting and changing with the needs of the user, robot and task.

One idea has been to couple the virtual world tightly with the robot knowledge of the world. The robot currently has access to the objects in the database and could in some way use these for its task needs. However in a situation where the model contains much more detail than is displayed at a given time, there may then be a need for another type of representation for the objects the robot can manipulate, navigate around, etc. Thus at some point the world may cease to be a just repository of robot knowledge. This balance of human user needs (e.g. to see model detail) and robot awareness (the objects the robot can act on) may be in tension with each other and need to be treated delicately. Different methods for doing this were considered in the design chapter.

7.1.2 Semi-autonomy design

It is intended that the competence of the robot can grown in this framework without major changes to the guiding metaphor of control. That is to say that as the robot can perform more autonomously, the requirements for changing the interface are proportional. Semi-autonomy could be seen as compound forms of previous behaviors where new behaviors are generated by combining old ones. The need on the interface side is then to aggregate the commands into new compound actions. This can be seen as an assistant becoming more competent over time. A number of respondents in the study asked for methods of imitation. That is to say, "I will show you how to do this task once, and then want you repeat it when I ask." This is in fact the way many factory robots perform their work. They are taught a sequence of moves using a 'teach pendant' and then set about to perform their task. This is similar also to the way a supervisor and trainee assistant interact. Thus such a behavior on the part of the robot would fit nicely into the greater scheme. Building flexibility, e.g. generalized taught plans, into such learned operations would form a strand of research on its own. This of course has been addressed in AI before. The hope here is that contextualizing the use might restrict the possibly paths of that generalization where it is not "intelligence" that is sought, but practicality

and usefulness.

7.2 Environment Visualization

In presenting the different methods for visualization a system that implements a concept called Augmented Virtuality has been demonstrated. The system roams around a real space while sending texture updates based on video to a virtual world model of the that space. Also shown were a number of other methods of displaying the video in the virtual environment. As demonstrated in the design study, some of these methods cause confusion and no one method is perfect for all situations.

Using the virtual model, a user can perform an off-line tour of the remote space in a form of tele-exploration. Using this application the user can decouple his actions from the actions of the robot in the remote space. Future work on reality portals centers on improving the quality of the automatic extraction by improving the initial and dynamic calibration of the camera as well as adding methods for automatic detection of objects by using simple image processing routines.

Another extension to these techniques could likewise be in worlds where no model has been supplied. In such situations a increased reliance on sensors and their ability to generate models of the world would need to be established. An simple example would be to use range sensors to sense solid objects. When such obstacles are detected a object could be instantiated in the location with an RP that extracts the texture. This information would prove useful to the human operator for identifying the object that triggered the response. A more sophisticated sensor system could also be constructed with basic image processing techniques coupled to the camera and calibration process. In this case, some structure, e.g. edges, could be determined from the scene, possibly with the user's assistance and then placed in the scene. Some sensors, such as laser range scanners, would be well suited to such a task working in parallel with a video camera. One can imagine a system composed of such a coupled range scanner and camera that quickly builds up a model of the environment complete with textures.

7.3 Robot Control and Interaction

From the design study it emerged that although users are willing to interact with the robot on its terms and competence level, they would prefer to interact at the highest level possible. Also of note is the fragility of the way the interface is presented. An example of that is presenting the robot's current state

and intentions. Users tended to became frustrated by longer pauses in robot response, even if the robot was operating as usual. In such cases a visual signal is required to inform the user that the robot is processing the last command and merely needs more time. Through action or visual signal the robot must indicate its current state.

Raising the level of control, the methods of specification, is of coursed linked to the abilities of the robot to perform. In most cases the user would like to merely issue a command. A good target level for such a command would be at the level of "go get my slippers." A device such as this would prove extremely useful and perhaps quite popular.

7.4 Experience of Building a 'User' System

The challenge of building a system to be used for extended periods at particular rigidly scheduled times should not be underestimated. It is valuable experience that puts new constraints and pressures on the system. In addition the experiences of integrating the sub-systems on a robot that *must* functioning in the real world with an understandable interface at a particular time, have contributed to the betterment of the system and to the knowledge of how to make such systems work in practice. These posed many challenges that were not predicted from the start. In the study also many new issues emerged that escaped prior observation through being either too close or too familiar with the system. Confounding the challenge was the issue of distance between the robot and the user position in the Study of Use. Ad hoc methods had to be developed to administer breakdowns using the available controls.

7.5 Study of Use

The study methodology employed in this thesis is primarily oriented toward discovering the specific, the exceptional, and the contingent within the context of use. The goal is in changing and improving the way the interface may be used to perform task-oriented collaborative work with the robot. The focus was on relevance, sometimes favoring specific ideas over general principles. One of the best approaches for this is involving users in an interactive process of systems design. This study should then not be seen as a final step in the design of a system, but rather a primary step in a long quest toward improvement through user and task centered design.

Study participants provided information that was examined for implications that suggest explicitly, or implicitly, ideas to be designed and engineered. Some of these ideas may be subtle and trivial to implement, others may involve restructuring of the system. Such a study also seeks to understand the implementation and ground the generation of new ideas. Working with users

of a system on a particular task within a study of use is one way to expose the problems and issues for further exploration. Such identification can then ground problem solving, further ideation and design iterations. Through a "tightness" with the users, the system designer can have a deeper understanding of both the potential problems and solutions at hand. It is in this way, through such direct contact with users, that engineering can be seen as an application oriented discipline that serves eminently practical purposes largely through systems design.

7.6 Final Words

As we enter the new millennium, it is clear that the trend of ubiquitous computing machines will increase. It is also clear that we are only seeing the beginning of a general launch of mobile computers with autonomous or semi-autonomous competence. These robots may become part of our daily lives. If indeed it is we that will control them, and it is they will help improve our quality of living, then a direction of potential application and user understanding is needed. There is a great deal of work open for exploring the potential applications and use of such systems in the home and office. Such work should be carried out in the context of user-centered research. Currently there is a great understanding of how to make more robust mobile robotic systems. Both potential users and robotics researchers will benefit by being brought together.

Survey Questionaire

This is a list of all the questions from the survey. The keywords in parenthesis after each question are the indices to the plots in Chapter 5.

Survey: Part I - before task

- What is your profession?
- If you do research, what is your research focus?
- How much computer competence would you say you have? (Computer competence)
- How much Robot experience have you had? (Robot experience)
- How familiar are you with remote space? (Remote familiarity)
- How much experience with virtual reality or 3D video games do you have as user? (VR familiarity)
- How would you rate your spatial ability in the physical world? (Spatial sense)
- How would you rate your visual ability in the physical world? (Visual ability)
- Would you say you think spatially? (Think spatially)
- Are you good with maps? (Map comfort)

- How familiar are you with the concept of telerobotics? (Telerobotics)
- How comfortable are you with computers? (Comfort with computers)
- Do you have trust in machine automation, e.g. factory automation, cash-machines, ..? (Machine trust)
- Can you imagine yourself employing a robot butler? (Robot butler)
- Are you comfortable delegating work in general? (Task Delegation)
- Do you enjoy being in charge? (In charge)
- Do you think the captain should go down with the ship? (Captain)
- What is your age, seniority (e.g. boss, student, etc)?

Survey: Part II - after task

- How effective was the Real vs. Virtual world correspondence? (Real vs. virtual)
- How strong was the sense of "Presence" in the remote physical space? (Presence remote)
- How strong was the sense of "Presence" in the virtual space? (Presence VR)
- Did you develop a Spatial feeling for the remote space (e.g. a sense of know where things are)? (Spatial feeling)
- Can your order from strongest to weakest the spatial sense you got from the following (e.g. 1 - 5):
  _graphical virtual world
  _textures in graphical world
  _video inside virtual world
  _from camera view
  _from robot's view
- Did the robot react to your commands? (Robot reaction)

- Was the robot assistive? (Robot assistive)
  To what extent did you feel you were the robot? And when? (Were robot)

- Did you understand what the robot was doing? (Understand robot)

- How much did you think the robot understood about its environment? (Robot understood)

- To what extent did you adapt or change your behavior for the robot? (Adapt to robot)

- Did the robot require your help? (Help robot)

- Was it difficult to find the boxes? (Task difficulty)

- Did you become frustrated? (Frustration)

- Did you have a sense of what you wanted to do? (Sense of what to do)

- Were you satisfied with the completion of the task? (Task satisfaction)

- Did you see yourself in the virtual world (video)?

Bibliography

[1] Electrolux AB. Electrolux robot vaccum cleaner prototype. *WWW address: http://www3.electrolux.se/robot/meny.html, 2/18/2000*, 1999.

[2] Julie A. Adams. Human factors experimental analysis of the masc system. Technical Report MS-CIS-96-11, University of Pennsylvania, March 1996.

[3] Julie A. Adams and Richard P. Paul. Experimental analysis of the mediation hierarchy theory. In *IEEE International Conference on Systems, Man and Cybernetics, Beijing, China*, 1996.

[4] Philip E. Agre. *Computation and Human Experience*. Cambridge University Press, 1997.

[5] Karl-Peter Åkesson. Augemented virtuality: A method to automatically augment virtual worlds with video images. Master's thesis, Chalmers University of Technology also Swedish Institute of Computer Science Technical Report, November 1997.

[6] Karl-Petter Åkesson and Kristian T. Simsarian. Reality portals. In *VRST 1999: Proceedings of Virtual Reality Software and Technology Symposium*, London, 1997.

[7] M. Andersson, A. Öreback, M. Lindström, and H.I. Christensen. *Intelligent Sensor Based Robots*, volume 1724 of *Lecture Notes in Artificial Intelligence*, chapter ISR: An Intelligent Service Robot, pages 291–314. Springer Verlag, Heidelberg, October 1999.

[8] Ronald Arkin. Reactive control as a substrate for telerobotic systems. *IEEE AES Systems Magazine*, pages 24–31, June 1991.

[9] Ronald Arkin and Khaled S. Ali. Integration of reactive and telerobotic control in multi-agent robotic systems. In *Proceedings of Third International Conference on Simulation of Adaptive Behavior: From Animals to Animats*, Brighten, UK, 1994.

[10] W.A. Aviles, T.W. Hughes, H.R. Everett, A.Y. Martin, and A.H. Koyamatsu. Issues in mobile robotics: The unmanned ground vehicle program teleoperated vehicle (TOV). In *SPIE Vol. 1388 Mobile Robots V*, pages 587–597, 1990.

[11] Dana H. Ballard. Animate vision. *Artificial Intelligence*, 48:57–86, 1991.

[12] John E. Bares and David S. Wettergreen. Lessons from the development and deployment of dante II. In *Proceedings of the 1997 Field and Service Robotics Conference*, Canberra, Australia, 1997.

[13] Ben B. Bederson and Angela Boltman. Does animation help users build mental maps of spatial information? In *Proceedings of Information Visualization Symposium (InfoVis 99)*, pages 28–35, New York, 1999. IEEE Press.

[14] S. Benford, B. Bederson, K. Åkesson, V. Bayon, A. Druin, P. Hansson, J. P. Hourcade, R. Ingram, H. Neale, C. O'Malley, K. Simsarian, D. Stanton, Y. Sundblad, and G.Taxén. Designing storytelling technologies to encourage collaboration between young children. In *CHI2000, ACM conference on Computer Human Interaction*. ACM Press, 2000.

[15] S. Benford, J. Bowers, L. Fahlén, and C. Greenhalgh. Managing mutual awareness in collaborative virtual environments. In *Proceedings of ACM VRST'94*, Singapore, 1994.

[16] S. Benford and L. Fahlén. A spatial model of interaction in large virtual environments. In *Third European Conference on Computer-Supported Cooperative Work*, pages 109–124. Kluwer Academic Publishers, 1993.

[17] Steve Benford, John Bowers, Lennart E. Fahlén, Chris Greenhalgh, John Mariani, and Tom Rodden. Networked virtual reality and co-operative work. *Presence*, 4(4), 1995.

[18] Steve Benford, Chris Greenhalgh, Gail Reynard, Chris Brown, and Boriana Koleva. Understanding and constructing shared spaces with mixed reality boundaries. *ACM Transaction on Computer-Human Interaction (ToCHI)*, 5(3), 1998.

[19] R. Bentley, J.A. Hughes, D. Randall, T. Rodden, P. Sawyer, and D. Shapiro I. Sommerville. Ethnographically-informed systems design for air traffic control. In *Conference on Computer-Supported Cooperative work, CSCW'92*, Toronto, 1992.

[20] Ted Blackmon. MarsMap - VR for Mars pathfinder. *WWW address: http://img.arc.nasa.gov/~blackmon/MarsMap/www_main.html, 10/2/1999*, 1998.

[21] Theodore T. Blackmon, F. Lai, and Lawrence W. Stark. Comparison of control modes during task learning with a telerobotics interface. *Automatica, to be published*, 1998.

[22] Theodore T. Blackmon and Lawrence W. Stark. Model-based supervisory control in telerobotics. *Presence*, 5(2):205–223, 1996.

[23] S. Bödker, P. Ehn, J. Kammersgaard, M. Kyng, and Y. Sundblad. A utopian experience. In G. Bjerknes, editor, *Computers and Democracy a Scandinavian Challenge*. Aldershoot, UK, 1987.

[24] S. Bouffouix and M. Bogaert. Real time navigation and obstacle avoidance for teleoperated vehicles. In *SPIE Vol. 1831 Mobile Robots VII*, pages 265–275, 1992.

[25] John Bowers. Making it work: A field study of a 'CSCW Network'. *The Information Society*, 11(3), 1995.

[26] Ivan Bretan. *Natural Language in Model World Interfaces.* Licentiate Thesis, Department of Computer and Systems Sciences. The Royal Institute of Technology and Stockholm University, Stockholm, Sweden, 1995.

[27] Rodney A. Brooks. Knowledge without reason. *MIT AI Tecnical Report*, number 1293, April 1993.

[28] W. Burgard, A.B. Cremers, D. Fox, D. Hähnel, G. Lakemeyer, D. Schulz, W. Steiner, and S. Thrun. Experiences with an interactive museum tour-guide robot. *Artificial Intelligence*, 114(1-2):3–55, 1999.

[29] N. Burtnyk and M.A. Greenspan. Supervised autonomy: partitioning telerobotic responsibilities between human and machine. In *International conference on intelligent teleoperation*, November 1991.

[30] Christer Carlsson and Olof Hagsand. DIVE – a platform for multi-user virtual environments. *Computers and Graphics*, 17(6), 1993.

[31] C. Chen and M.M. Trivedi. Savic: A simulation, visualization, and interactive control environment for mobile robots. *International Journal of Pattern Recognition and Artificial Intelligence*, 7:123–144, 1993.

[32] Sony Coporation. Sony AIBO ERS-111, robotic pet dog. *WWW address: http://www.world.sony.com/Electronics/aibo/index.html, 2/18/2000*, 1999.

[33] John J. Craig. *Introduction to Robotics: Mechanics and Control.* Addison-Wesley, 1989.

[34] Brady R. Davies, Michael J. McDonald, and Raymond W. Harrigan. Virtual collaborative environments: Programming and controlling robotic devices remotely. In *SPIE Proceedings, Telemanipulator and Telepresence Technologies*, Boston, October 1994.

[35] Gregory A. Dorais, R. Peter Bonasso, David Kortenkamp, Barney Pell, and Debra Schreckenghost. Adjustable autonomy for human-centered autonomous systems. In *Sixteenth International Joint Conference on Artificial Intelligence Workshop on Adjustable Autonomy Systems*, Stockholm, 1999.

[36] Allison Druin. Cooperative inquiry: Developing new technologies for children with children. In *CHI99, ACM conference on Computer Human Interaction.* ACM Press, 1999.

[37] Jeffrey M. Bradshaw (ed.). *Software Agents.* MIT Press, 1997.

[38] D. England, W. Prinz, K.T. Simsarian, and O. Ståhl. A virtual environment for collaborative administration. In *IEEE International Conference on Virtual Environments on the Internet, WWW and Network*, New York, April 1997.

[39] Steven Feiner, Blair MacIntyre, and Doree Seligmann. Knowledge-based augmented reality. *Communications of the ACM*, 36(7):52–62, July 1993.

[40] Cornelia Fermüller and Yiannis Aloimonos. Vision and action. *Image and Computing*, 13(10), 1995.

[41] W.R. Ferrel and T.B. Sheridan. Supervisory control of remote manipulation. *IEEE Spectrum*, 4(10):81–88, Oct 1967.

[42] Richard Feynman. *The meaning of it all.* Penguin Books, 1998.

[43] Richard E. Fikes, Peter E. Hart, and Nils J. Nilsson. Learning and executing generalized robot plans. *Artificial Intelligence*, 3:251–288, 1972.

[44] S. S. Fisher, M. Mcgreevy, J. Humphries, and W. Robinet. Virtual interface environments for telepresence applications. In J.D. Berger, editor, *Proceedings of ANS International Topical Meeting of Remote Systems and Robotics in Hostile Environments*, 1987.

[45] Scott Fisher. Virtual interface environments. In Brenda Laurel, editor, *The Art of Human-Computer Interface Design.* Addison-Wesley, 1990.

[46] L. Foner. What's an agent anyway? A sociological case study. Technical report, Agents Memo, Media Lab, MIT, 93-01 1993.

[47] D. Fox, W. Burgard, and S. Thrun. Dynamic window approach to collision avoidance. *IEEE Robotics and Automation*, 4(1), 1997.

[48] Jannik Fritsch, Hans Brandt-Pook, and Gerhard Sagerer. Combining planning and dialog for cooperative assembly construction. In *IJCAI-99 Workshop: Scheduling and Planning meet Real-time Monitoring in a dynamic and uncertain world*, 1999.

[49] Ricki Goldman-Segall. *Points of Viewing Children's Thinking: A Digital Ethnographer's journey.* Lawrence Erlbaum, New Jersey, 1998.

[50] Olof Hagsand. Interactive multiuser VEs in the DIVE system. *IEEE Multimedia*, 3(1), 1996.

[51] Pär Hansson, Anders Wallberg, and Kristian Simsarian. Techniques for 'Natural' interaction in multi-user CAVE-like environments. In *Short paper in Proceedings of ECSCW '97*, Lancaster, UK, 1997.

[52] Christian Heath, Marinna Jirotka, Paul Luff, and Jon Hindmarsh. Unpacking collaboration: the interactional organization of trading in a city dealing room. In *Proceedings of third European Conference on Computer Supported Cooperative work - ECSCW 93*, Milan, 1993.

[53] C. Heeter. Being there: the subjective experience of presence. *Presence*, 1(2):262–271, 1992.

[54] Butler Hine, Phil Hontalas, Terrence Fong, Laurent Piguet, Erik Nygren, and Aaron Kline. VEVI: a virtual environment teleoperations interface for planetary exploration. In *SAE 25th International Conference on Environmental Systems*, San Diego, 1995.

[55] Erik Hollnagel. AI+HCI: Much ado about nothing? In *Computer Resources International*, Copenhagen, July, 1989.

[56] Kristina Höök. Steps to take before intelligent user interfaces become real. *Journal of Interacting with Computers*, 12, 4, Feb. 2000.

[57] Ian Horswill. A simple, cheap, and robust visual navigation system. In *From Animals to Animats II: Proceedings of second international conference of adaptive behavior*, 1993.

[58] John A. Hughes and David Randalland Dan Shapiro. Faltering from ethnography to design. In *Conference on Computer-SUpported Cooperative work, CSCW'92*, Toronto, 1992.

[59] Nelson Corby Jr. and Christopher Nafis. Augmented reality telemanipulation system for nuclear reactor inspection. In *SPIE Proceedings, Telemanipulator and Telepresence Technologies*, Boston, October 1994.

[60] Jussi Karlgren, Ivan Bretan, Niklas Frost, and Lars Jonsson. Interaction models, reference, and interactivity for speech interfaces to virtual environments. *Proceedings of Second Eurographics Workshop on Virtual Environments – Realism and Real Time*, 1995.

[61] Arun Katkere, Saied Moezzi, Don Kuramura, Patrick Kelly, and Ramesh Jain. Towards video-based immersive environments. *Special Issue on Multimedia and Multisensory Virtual Worlds*, May 1996.

[62] Alan Kay. Computer software. *Scientific American*, 251(3), 1984.

[63] Z. Kazi, M. Beitler, M. Salaganicoff, S. Chen, D. Chester, and R. Foulds. Intelligent telerobotic assistant for people with disabilities. In *SPIE Proceedings, Telemanipulator and Telepresence Technologies II*, volume 2590, Philadelphia, October 1995.

[64] Z. Kazi, S. Chen, M. Beitler, D. Chester, and R. Foulds. Multimodally controlled intelligent telerobot for people with disabilities. In *SPIE Proceedings, Telemanipulator and Telepresence Technologies III*, volume 2901, Boston, November 1996.

[65] Jacqueline H. Kim, Richard J. Weidner, and Allan L. Sacks. Using virtual reality for science mission planning: A Mars pathfinder case. In *ISMCR 1994: Topical Workshop on Virtual Reality*, pages 37–42, Houston, 1994. NASA Conference publication 10163.

[66] Gudrun J. Klinker, Klaus H. Ahlers, David Breen, Pierre-Yves Chevalier, Chris Compton, Douglas Greer, Dieter Koller, Andres Kramer, Eric Rose, Mihran Tuceryan, and Ross T. Whitaker. Confluence of computer vision and interactive graphics for augmented reality. *Presence*, 6(4), August 1997.

[67] Myron Kreugar. *Artificial Reality II.* Addison-Wesley, 1991.

[68] Jaron Lanier and Pattie Maes. Intelligent agents = stupid humans? *Hotwired debate, http://www.hotwired.com/*, July 15-24, 1996.

[69] M. Li. Camera calibration of the s head-eye system. Technical report, Computational Vision and Active Perception Lab., Dept. of Numerical Analysis and Computing Science, Royal Institute of Technology (KTH), March 1994.

[70] Friendly Machines Ltd. Robomow, the intelligent autonomous lawnmower. *WWW address: http://www.friendlymachines.com/, 2/18/2000*, 1999.

[71] Douglas MacKenzie and Ronald Arkin. Evaluating the usability of robot programming toolsets. *The International Journal of Robotics Research*, 17(4), August 1987.

[72] P. Maes and B. Shneiderman. Direct manipulation vs. interface agents: a debate. *Interactions*, IV(6), 1997.

[73] M. Mallem, F. Chavand, and E. Colle. Computer-assisted visual perception in teleoperated robotics. *Robotica (10)*, pages 99–103, 1992.

[74] Drew McDermott. Artificial intelligence meets natural stupidity. *SIGART Newsletter*, 57, 1976.

[75] Paul Milgram and David Drascic. Enhancement of 3-d video displays by means of superimposed stereo-graphics. In *Proceedings of the Human Factors Society 35th Annual Meeting*, pages 1457–1461, 1991.

[76] Paul Milgram, Anu Rastogi, and Julius Grodski. Telerobotic control using augmented reality. In *Proceedings 4th IEEE Internation Workshop on Robot and Human Communication*, Tokyo, July 1995.

[77] Tom Mitchell. Lecture containing a new challenge to behavior-based robotics. In *NATO Advanced Study Institute: The Biology and Technology of Autonomous Agents*, 1993.

[78] NASA. *Proceedings of the NASA Conference on Space Telerobotics*. JPL Publication 89-7, Vol 1-5, Pasadena, Ca, 1989.

[79] E. Natonek, L. Flückiger, Th. Zimmerman, and C. Baur. Virtual reality: an intuitive approach to robotics. In *SPIE Proceedings, Telemanipulator and Telepresence Technologies*, Boston, October 1994.

[80] Jakob Nielsen. Guerrilla HCI: Using discount usability engineering to penetrate the intimidation barrier. In R. G. Bias and D. J. Mayhew, editors, *Cost-Justifying Usability*. Academic Press, 1994.

[81] Jakob Nielsen. *Usability Engineering*. Morgan Kaufmann Publishers, 1994.

[82] Don Norman. *The Invisible Computer*. MIT Press, 1998.

[83] Donald A. Norman. *The Design of Everyday Things*. Doubleday Books, 1990.

[84] Lars Oestreicher, Helge Huettenrauch, and Kerstin Severinsson-Eklundh. Where are you going little robot? - prospects of human-robot interaction. In *Basic Research Symposium of CHI99, ACM conference on Computer Human Interaction*, 1999.

[85] PersonalRobotics.com. The Cye vacuum robot. *WWW address: http://www.personalrobotics.com, 2/18/2000*, 1999.

[86] Polly K. Pook. *Teleassistance: Using Deitic Gestures to Control Robot Action*. PhD thesis, Department of Computer Science, College of Arts and Sciences, University of Rochester, New York, PhD Dissertation, May 1995.

[87] Eric Prem. Elements of an epistemology of embodied AI. In *AAAI Fall Symposium on Empodied Cognition and Action*, Menlo Park, CA, 1996.

[88] Peter Rander, P.J. Narayanan, and Takeo Kanade. Virtualized reality:constructing time-varying virtual worlds from real events. In *Proceedings of IEEE Visualization '97*, pages 277–283, Phoenix, Arizona, October 1997.

[89] Byron Reeves and Clifford Nass. *The Media Equation : How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press, 1996.

[90] IEEE ROMAN. *IEEE International Workshop on Robot and Human Communication*. IEEE Press, Biannual Conference.

[91] Phoebe Sengers. Do the thing right: an architecture for action-expression. In *Proceedings of the second international conference on Autonomous agents*, pages 24–31, Minneapolis, MN, May 10-13 1998.

[92] Thomas B. Sheridan. *Telerobotics, Automation and Human Supervisory Control*. MIT Press, Cambridge, MA, 1992.

[93] Donna Shirley. *Managing Martians*. Broadway Books, 1999.

[94] Ben Shneiderman. Natural vs. precise concise languages for human operation of computers: Research issues and experimental approaches. In *Proceedings of the 18th Annual Meeting of the Association for Computational Linguistics*, Philadelphia, 1980.

[95] Ben Shneiderman. *Designing the User Interface : Strategies for Effective Human-Computer Interaction*. Addison-Wesley Publishers, 1997.

[96] Ben Shneiderman. Direct manipulation versus agents: Paths to predictable, controllable, and comprehensive interfaces. In Jeffrey Bradhshaw, editor, *Software Agents*, pages 97–106, 1997.

[97] Kristian Simsarian. Multi-modal active sensing for a simple mobile agent. In *NATO Advanced Study Institute: The Biology and Technology of Autonomous Agents*. Springer Verlag, 1993.

[98] Kristian T. Simsarian. View-invariant regions and mobile robot self-localization. Master's thesis, University of Virginia, May 1991.

[99] Kristian T. Simsarian. A system for robotic telepresence employing VR as the communication medium: Interface metaphors. In Henrik I. Christensen, Carsten Bräutigam, and Christian Ridderström, editors, *Proceedings of the 5th international Symposium on Intelligent Robotic Systems 1997*, pages 5–11, 1997.

[100] Kristian T. Simsarian and Karl-Petter Åkesson. Windows on the world: An example of augmented virtuality. In *Interface 1997, Sixth International Conference Montpellier 1997: Man-machine interaction*, pages 68–71, 1997.

[101] Kristian T. Simsarian and L. Fahlén. Using virtual and augmented reality to control an assistive mobile robot. In *Proceedings of Virtual Reality and Persons with Disabilities*, San Francisco, Aug 1995.

[102] Kristian T. Simsarian, L. Fahlén, and Emmanuel Frécon. Virtually telling robots what to do. In *Fourth International Conference of Informatique 1995, Montpellier France, Interface to real and virtual worlds*, pages 511–520, 1995.

[103] Kristian T. Simsarian, Jussi Karlgren, L. Fahlén, Emmanuel Frécon, Ivan Bretan, Niklas Frost, Lars Jonsson, and Tomas Axling. Achieving virtual presence with a semi-autonomous robot through a multi-reality and speech control interface. In M. Goebel, J. David, P. Slavik, and J.J. van Wijk, editors, *Virtual Environments and Scientific Visualization '96*. SpringerCS, 1996.

[104] Kristian T. Simsarian, Thomas J. Olson, and N. Nandhakumar. View-invariant regions and mobile robot self-localization. *IEEE Transactions on Robotics and Automation*, Oct 1996.

[105] Mel Slater and Sylvia Wilbur. A framework for immersive virtual environments (five): Speculations on the role of presence in virtual environments. *Presence*, 6(6), 1998.

[106] Tim Smithers. Taking eliminative materialism seriously: A methodology for autonomous systems research. In *NATO Advanced Study Institute: The Biology and Technology of Autonomous Agents*, 1993.

[107] Lawrence W. Stark and Yun S. Choi. Experimental metaphysics: The scanpath as an epistomological mechanism. In W.H. Zangemeister, H.S. Stiehl, and C. Freksa, editors, *Visual attention and cognition*. Elsevier Science B.V., 1996.

[108] Luc Steels. Building agents out of autonomous behavior systems. In L. Steels and R. Brooks, editors, *The Artificial Life Route to Artificial Intelligence*. Lawerence Erlbaum, 1993.

[109] Lucy Suchman. *Plans and Situated actions: The problems of Human-Machine communication*. Cambridge University press, 1987.

[110] Yngve Sundblad and Olle Sundblad. Olga - a multimodal interactive information assistant. In *CHI98, ACM conference on Computer Human Interaction*, Los Angeles, 1998. ACM Press.

[111] Gecko Systems. The CareBot PCR. *WWW address: http://www.geckosystems.com/, 2/19/2000*, 2000.

[112] Kristinn R. Thrisson. *Communicative Humanoids: A Computational Model of Psychosocial Dialogue*. PhD thesis, Submitted to the Program in Media Arts and Sciences, School of Architecture and Planning, Massachusetts Institute of Technology, July 1996.

[113] R. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of robotics and automation*, RA-3(4):323–344, August 1987.

[114] Sherry Turkle. *The Second Self: Computers and the human spirit*. Simon and Shuster, New York, 1984.

[115] Tomas Uhlin. *Fixation and seeing system*. PhD thesis, Computer Vision and active perception (CVAP), Royal Institute of Technology, Stockholm, Sweden, PhD Dissertation, May 1996.

[116] Carnegie Mellon Universirty and RedZone. Pioneer, the Chernobyl remote reconnaissance system. *WWW address: http://www.frc.ri.cmu.edu/projects/pioneer/, 2/23/2000*, 1999.

[117] Mark Weiser. The computer for the twenty-first century. *Scientific American*, Sept, 1991.

[118] Robert Stuart Weiss. *Learning from Strangers: The Art and Method of Qualitative Interview Studies*. Free Press, 1995.

[119] Robert B. Welch, Theodore T. Blackmon, Andrew Liu, Barbara A. Mellers, and Lawrence W. Stark. The effects of pictorial realism, delay of visual feedback, and observer interactivity on the subjective sense of presence. *Presence*, 5(3):263–273, 1996.